Struktury Baz Danych

Miłosz Ilecki

184577

Spis treści **1**

Wstęp **2**

Opis metody **3**

Typ rekordu **3**

Eksperyment **4**

Dane **4**

Wykresy **5-7**

Komentarz do eksprymentu **8**

Wniosek **9**

Wstęp

Projekt polegał napisaniu program sortującego plik metodą scalania naturalnego. Taśmy powinny być zrealizowane w formie plików dyskowych. Należy zaimplementować symulację odczytu oraz zapisu blokowego jako oddzielną warstwę logiki programu. Program powinien wyświetlić zawartość pliku przed sortowaniem i po posortowaniu. Program powinien też dawać możliwość wyświetlenia pliku po każdej fazie sortowania. Na zakończenie należy wyświetlić liczbę faz sortowania oraz liczbę odczytów i zapisów stron na dysk. Dodatkowo program powinien dawać możliwość wczytywania danych testowych z pliku testowego.

.

Podczas implementacji korzystałem z bibliotek:

- *math* - zaokrąglanie liczb, logarytmy, pierwiastki
- *copy*- tworzenie kopii obiektów,
- *random*- generowanie losowych danych,

Opis Metody

W tym projekcie korzystałem z metody Natural merge 2+1. Cechowała się ona trzema ilościami taśm. Na początku na taśmie nr 1, generowałem losowe dane o wybranej wielkości. Taśma nr 1 była następnie rozdzielana seriami pomiędzy taśmę nr 2 i 3. Ograniczając się wielkością bufora do odczytu/zapisu. Gdy zostały rozdzielone na dwie różne taśmy następował moment ich złączenia do taśmy nr 1, stosując sortowanie przez wstawianie między seriami. To określało jedną fazę. Algorytm kończył się gdy na taśmie nr 1 została posortowana lista.

Typ Rekordu

Na enauczaniu wylosowałem **13. Rekordy pliku: Liczby naturalne**

Specyfikacja pliku

Opisane wcześniej taśmy były reprezentowane za pomocą plików binarnych. Liczba naturalna zostawała zapisywana na 4 bajtach. Nie stosowano żadnych rozdzielników w celu oszczędności miejsca.

Przykładowo liczba: **645** przechowana w formacie bajtów **b'\x00\x00\x02\x85'**.

Prezentacja wyników

Program wypisuje posortowane dane wraz z ich wartościami.

liczby : **[13, 280, ….**

sumy : **[14, 720, ….**

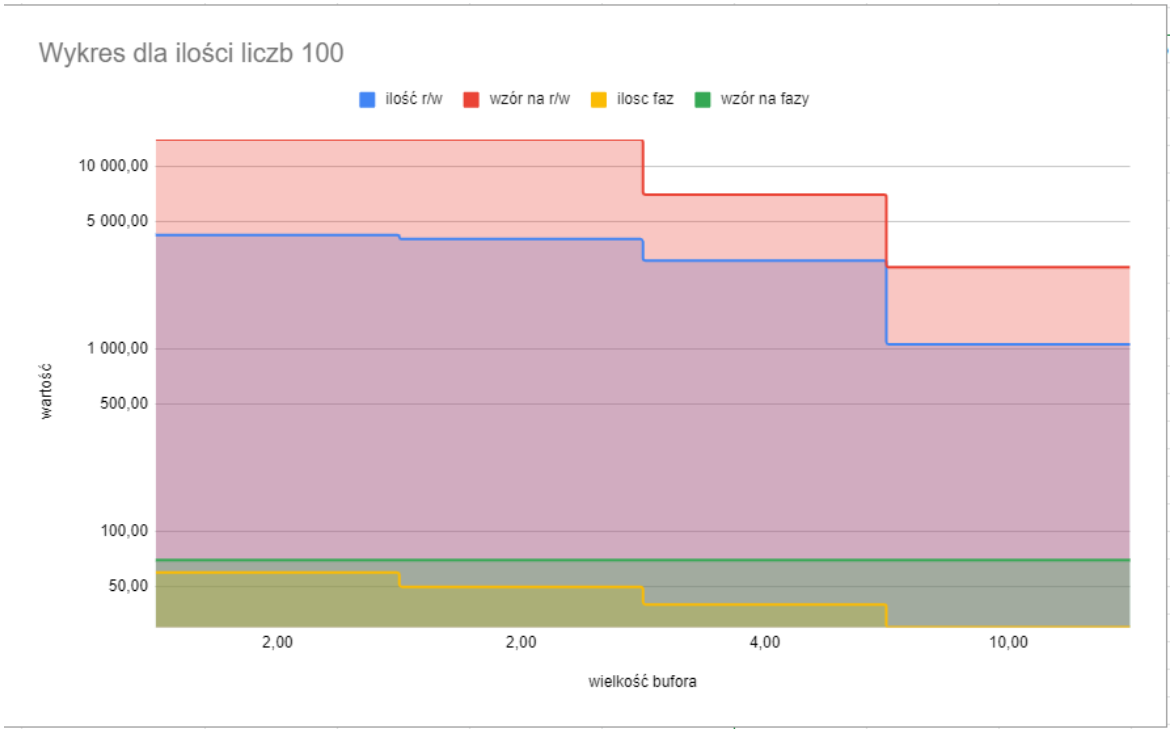
Gdzie liczby to kolejne posortowane liczby względem ich sum.

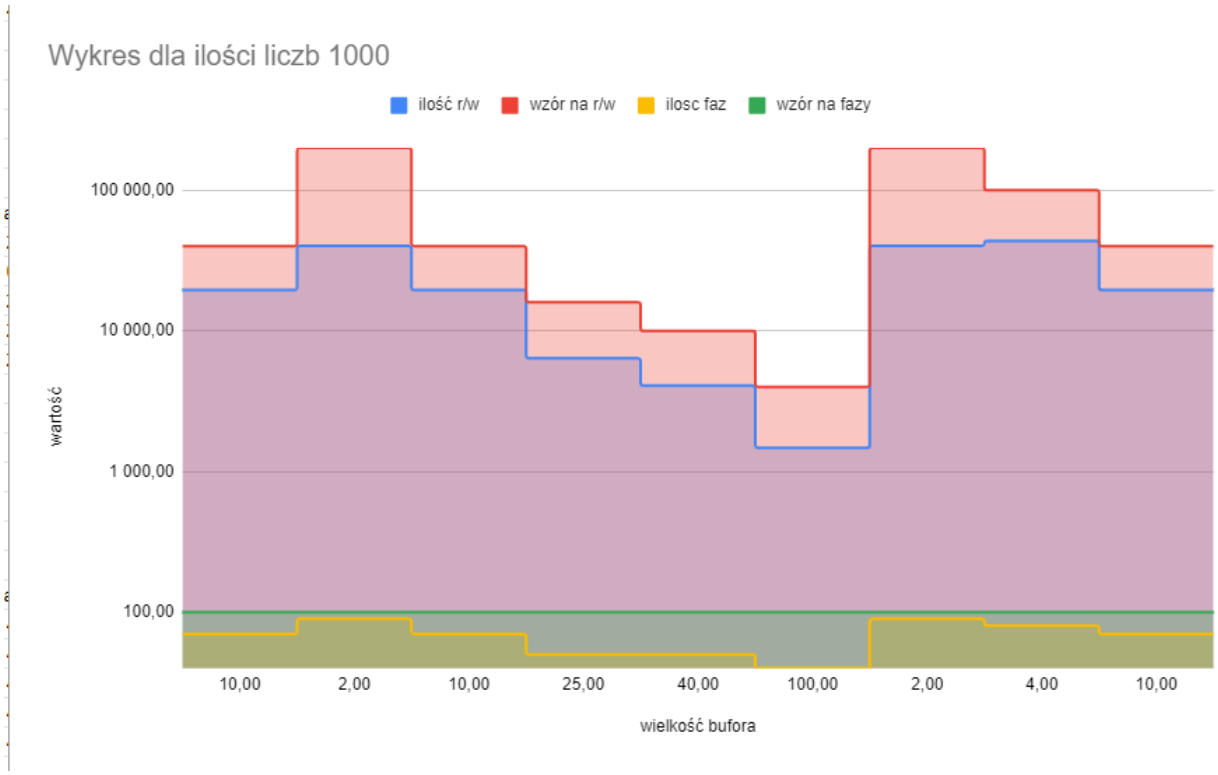
Eksperyment

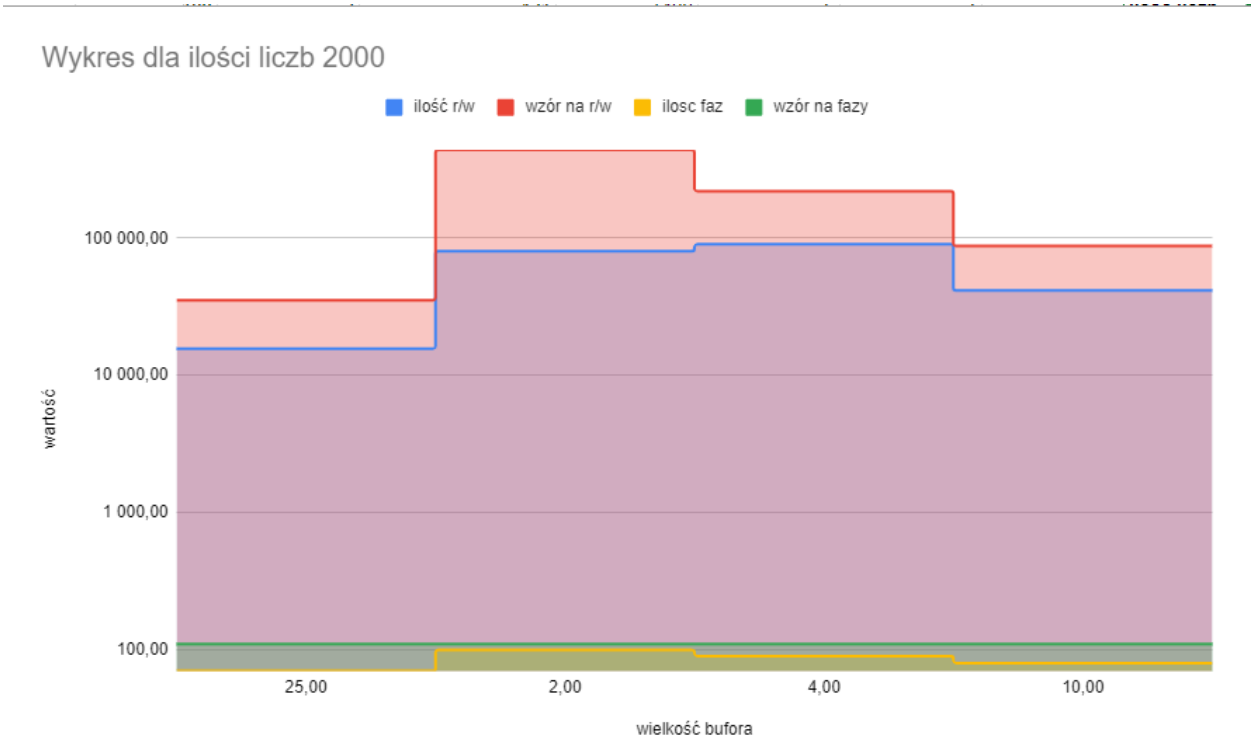
Stosowałem różne zmienne dane takie jak: wielkość pliku, wielkość bufora.

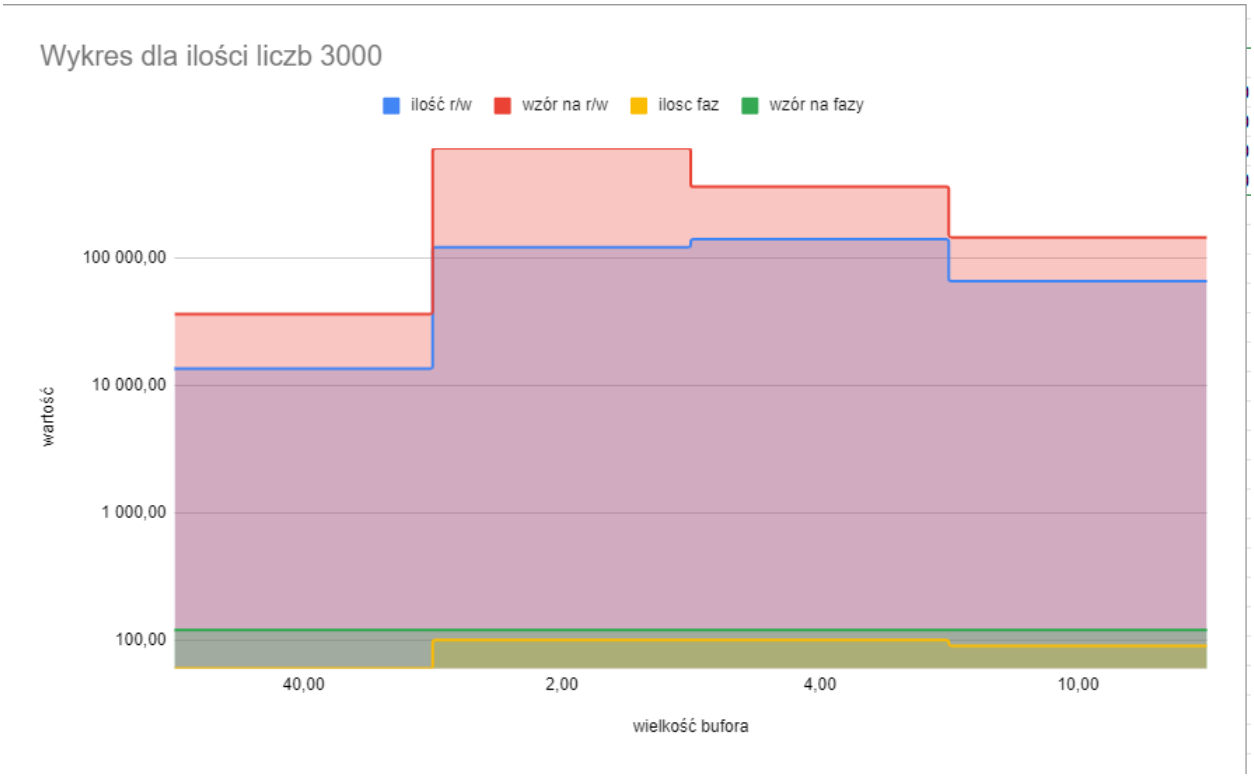
ilość liczb	wielkość bufora	ilość r/w	wzór na r/w	ilość faz	wzór na fazy
100,00	2,00	421,00	1 400,00	6,00	7,00
100,00	2,00	400,00	1 400,00	5,00	7,00
1 000,00	2,00	4 030,00	20 000,00	9,00	10,00
1 000,00	2,00	4 030,00	20 000,00	9,00	10,00
2 000,00	2,00	8 037,00	44 000,00	10,00	11,00
3 000,00	2,00	12 021,00	72 000,00	10,00	12,00
10 000,00	2,00	40 045,00	280 000,00	12,00	14,00
100,00	4,00	305,00	700,00	4,00	7,00
1 000,00	4,00	4 369,00	10 000,00	8,00	10,00
2 000,00	4,00	9 038,00	22 000,00	9,00	11,00
3 000,00	4,00	13 915,00	36 000,00	10,00	12,00
10 000,00	4,00	47 943,00	140 000,00	12,00	14,00
100,00	10,00	106,00	280,00	3,00	7,00
1 000,00	10,00	1 954,00	4 000,00	7,00	10,00
1 000,00	10,00	1 954,00	4 000,00	7,00	10,00
1 000,00	10,00	1 954,00	4 000,00	7,00	10,00
2 000,00	10,00	4 149,00	8 800,00	8,00	11,00
3 000,00	10,00	6 536,00	14 400,00	9,00	12,00
10 000,00	10,00	22 563,00	56 000,00	10,00	14,00
1 000,00	25,00	638,00	1 600,00	5,00	10,00
2 000,00	25,00	1 568,00	3 520,00	7,00	11,00
1 000,00	40,00	408,00	1 000,00	5,00	10,00
3 000,00	40,00	1 341,00	3 600,00	6,00	12,00
1 000,00	100,00	147,00	400,00	4,00	10,00
10 000,00	100,00	1 954,00	5 600,00	7,00	14,00

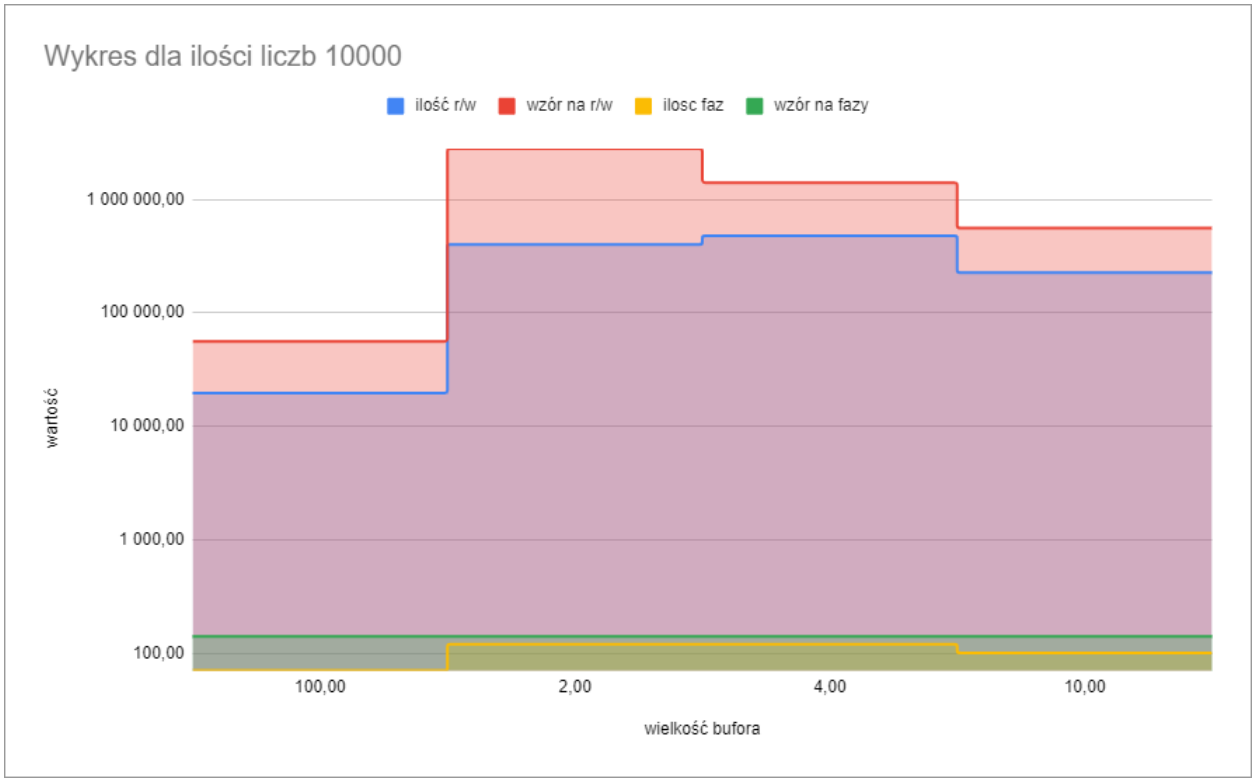
Na bazie tych rekordów stworzyłem wykresy.

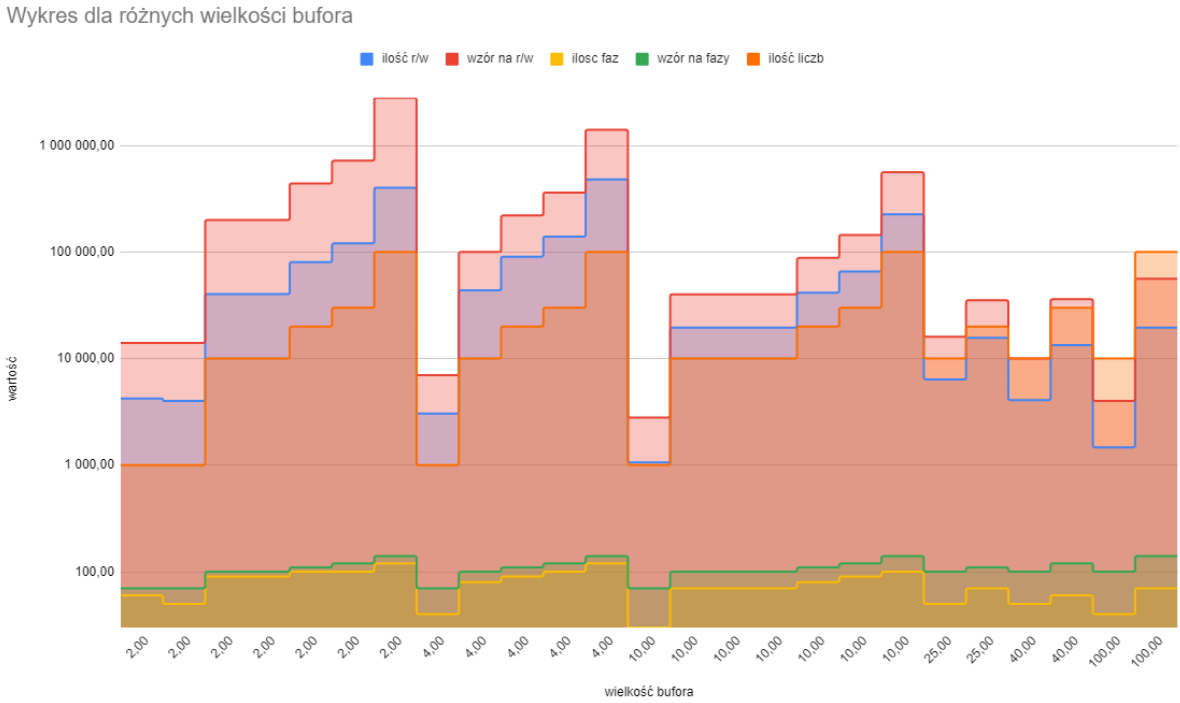

Wykres dla ilości liczb 100
ilość r/w
wzór na r/w
ilosc faz
wzór na fazy
10 000,00
5 000,00
1 000,00
500,00
100,00
50,00
wartość
2,00
2,00
4,00
10,00
wielkość bufora


Wykres dla ilości liczb 1000
ilość r/w
wzór na r/w
ilosc faz
wzór na fazy
100 000,00
10 000,00
1 000,00
100,00
wartość
10,00
2,00
10,00
25,00
40,00
100,00
2,00
4,00
10,00
wielkość bufora


Wykres dla ilości liczb 2000
ilość r/w
wzór na r/w
ilosc faz
wzór na fazy
100 000,00
10 000,00
1 000,00
100,00
wartość
25,00
2,00
4,00
10,00
wielkość bufora


Wykres dla ilości liczb 3000
ilość r/w
wzór na r/w
ilosc faz
wzór na fazy
100 000,00
10 000,00
1 000,00
100,00
wartość
40,00
2,00
4,00
10,00
wielkość bufora


Wykres dla ilości liczb 10000
ilość r/w
wzór na r/w
ilosc faz
wzór na fazy
1 000 000,00
100 000,00
10 000,00
1 000,00
100,00
wartość
100,00
2,00
4,00
10,00
wielkość bufora


Wykres dla różnych wielkości bufora
ilość r/w
wzór na r/w
ilosc faz
wzór na fazy
ilość liczb
1 000 000,00
100 000,00
10 000,00
1 000,00
100,00
wartość
2,00
4,00
10,00
25,00
40,00
100,00
wielkość bufora

Z pierwszych pierwszych wykresów można wizualnie zobaczyć jak wpływają one na inne wyniki.

Ostatni wykres przedstawia zmienną wielkość bufora.

Do wykresu wliczają się wartości porównawcze były one liczone ze wzorów:

log2 N - ilość faz

4*N log2 N / b - ilość odczytów i zapisów

Są to wzory to oszacowań górnych, czyli najgorszych przypadków, dlatego nasze rekordy nie będą ich przekraczać.

Wniosek

Sortowanie metodą naturalną ma swoje korzyści poprzez małą pamięć operacyjną możemy posortować większe dane objętościowo. Jest to wydajny algorytm.

Obecne wyniki w tabeli lub na wykresach łapią się pod najgorszym scenariuszem.

Dane były pseudo-losowe co daje możliwość rozbieżnych wyników.