

# STRUKTURY BAZ DANYCH

## Zadanie 2

### 1. WSTĘP

W ramach projektu zaimplementowałem strukturę B-drzewa. Każdą operację tj. wstawianie, usuwanie oraz aktualizację poprzedzam operacją wyszukiwania, która znajduje w indeksie odpowiedni klucz lub idealne dla niego miejsce. Jednocześnie operacja ta zapisuje przeczytane strony do bufora zaimplementowanego w postaci listy. W indeksie strona będąca korzeniem znajduje się zawsze na początku, a pozostałe strony zgodnie z kolejnością ich tworzenia. Utworzyłem również mechanizm, który przechowuje wszystkie zwolnione strony i miejsca w pliku z rekordami w osobnych plikach w postaci 4B adresu tj. miejsce w odpowiednim pliku.

Strony w indeksie prócz kluczy, adresów do stron potomnych oraz rekordów w nagłówku posiadają również 4B pole z informacją o liczbie kluczy przechowywanych na stronie.

Program został napisany w języku Java.

#### **Rekordy pliku:**

prędkość (float 4B) i masa obiektu (double 8B).

Łącznie każdy rekord zajmuje 12B. Strony w indeksie, w pliku z rekordami oraz w plikach pomocniczych mają tę samą wielkość strony obliczaną na podstawie **d**.

### 2. DOSTĘPNE KOMENDY

Programem steruje się za pomocą komend, oto przykłady:

```
init -i index.txt -r records.txt -d 10  
insert -k 16 -p i  
remove -k 16 -p r  
update -k 16 -p n  
read -k 16  
exit
```

parametr p odpowiada za sposób prezentowania struktury po danej operacji. Możliwe opcje:

- i (wyświetlany są wszystkie klucze z indeksu)
- r (wyświetlane są wszystkie rekordy z pliku z rekordami)
- n (nic nie jest wyświetlane)

Program sam losuje prędkość i masę obiektu.

### 3. OPIS EKSPERYMENTU

Eksperyment polegał na wstawieniu do drzewa określonej liczby kluczy, a następnie na wykonaniu odpowiedniej operacji w ilości: 10% liczby rekordów. Następnie obliczyłem średnią liczbęostępów do plików.

D=6

Liczba rekordów	Wstawianie		Usuwanie		Aktualizacja		Wart. Teor.
	Wart. śr.	Bł. wzgl.	Wart. śr.	Bł. wzgl.	Wart. śr.	Bł. wzgl.	Wart. śr.
100	5,50	113,99%	4,30	67,30%	3,90	51,74%	2,57
1000	7,21	87,02%	4,20	8,94%	4,89	26,84%	3,86
10000	8,05	56,53%	5,43	5,63%	5,91	14,97%	5,14
100000	9,08	41,39%	6,59	2,56%	6,92	7,65%	6,43
1000000	10,02	29,89%	7,83	1,55%	7,91	2,63%	7,71

D=12

Liczba rekordów	Wstawianie		Usuwanie		Aktualizacja		Wart. Teor.
	Wart. śr.	Bł. wzgl.	Wart. śr.	Bł. wzgl.	Wart. śr.	Bł. wzgl.	Wart. śr.
100	5,70	207,57%	3,50	88,86%	4,00	115,84%	1,85
1000	6,61	137,78%	4,13	48,57%	4,97	78,78%	2,78
10000	6,65	79,41%	4,07	9,73%	4,96	33,85%	3,71
100000	7,67	65,49%	5,05	9,03%	5,96	28,56%	4,63
1000000	8,71	56,58%	6,06	9,01%	6,96	25,14%	5,56

D=20

Liczba rekordów	Wstawianie		Usuwanie		Aktualizacja		Wart. Teor.
	Wart. śr.	Bł. wzgl.	Wart. śr.	Bł. wzgl.	Wart. śr.	Bł. wzgl.	Wart. śr.
100	6,00	290,31%	3,00	95,15%	4,00	160,21%	1,54
1000	5,46	136,79%	3,07	33,14%	3,97	72,17%	2,31
10000	6,47	110,31%	4,05	31,83%	4,97	61,78%	3,07
100000	7,48	94,50%	5,03	30,81%	5,98	55,53%	3,84
1000000	7,47	61,94%	5,03	8,97%	5,97	29,55%	4,61

Wartości teoretyczne obliczyłem ze wzoru  $\log_d N$ , gdzie N to liczba rekordów. Są to przybliżone średnie ilościostępów do pliku w procesie wyszukiwania. Otrzymane przez ze mnie wyniki nie są o wiele większe. Różnica wynika z faktu że operację wstawiania, usuwania czy aktualizacji są obarczone dodatkowymi dostępami do stron sąsiednich w procesie chociażby kompensacji, scalania czy rozszczepiania. Średnie ilościostępów po plików w procesie usuwania są troszeczkę mniejsze niż opowiadające im ilości w procesie aktualizacji. Spowodowane jest to tym, że nie zeruje 12B po usuniętym rekordzie.

Operację czytania całego indeksu wywołałem tylko raz i zanotowałem ilościostępów do pliku.

D=20

w	L. dost.	Max wart.
100	5	5,95
1000	30	50,95
10000	259	500,95
100000	2516	5000,95
1000000	25156	50000,95

Maksymalną wartość obliczyłem ze wzoru  $(N-1)/d + 1$ . Moje wyniki zgadzają się z wartościami teoretycznymi.

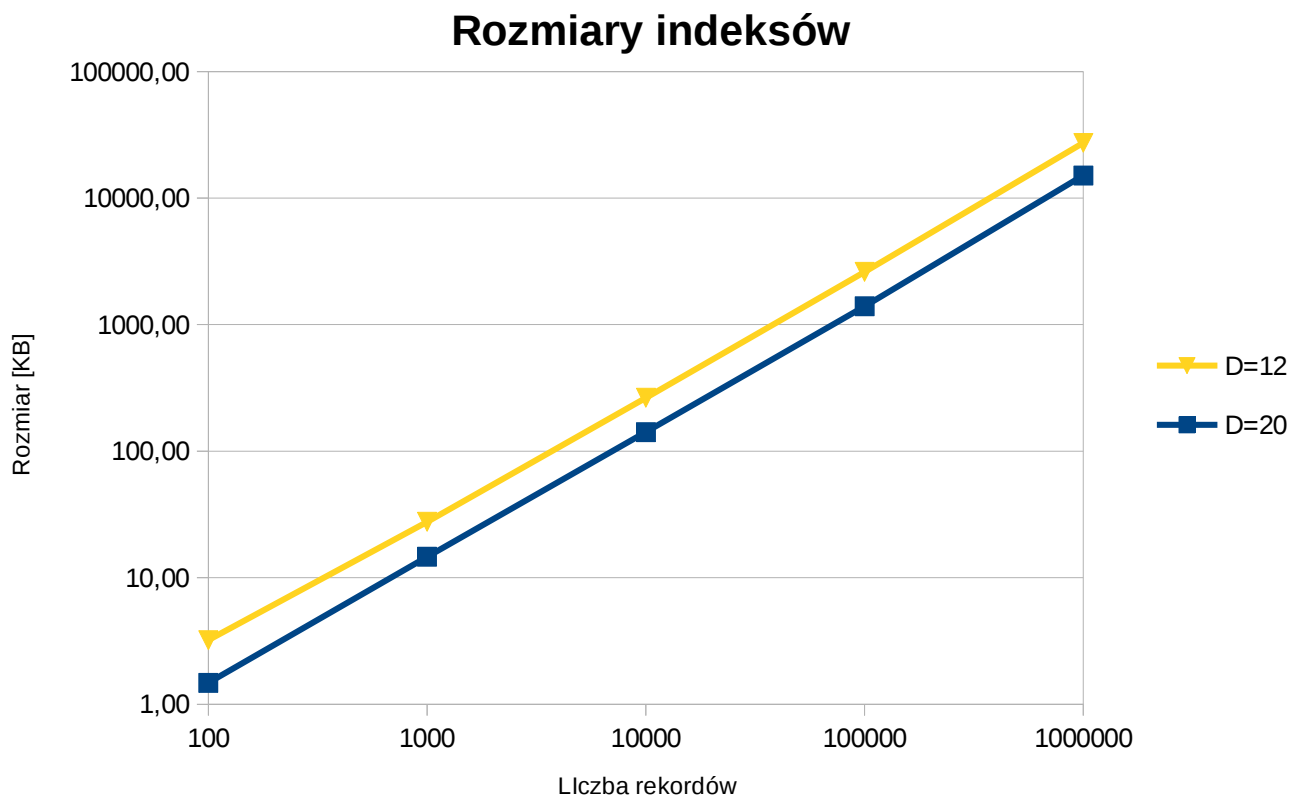
Na koniec postanowiłem sprawdzić rozmiary indeksu i pliku z rekordami.

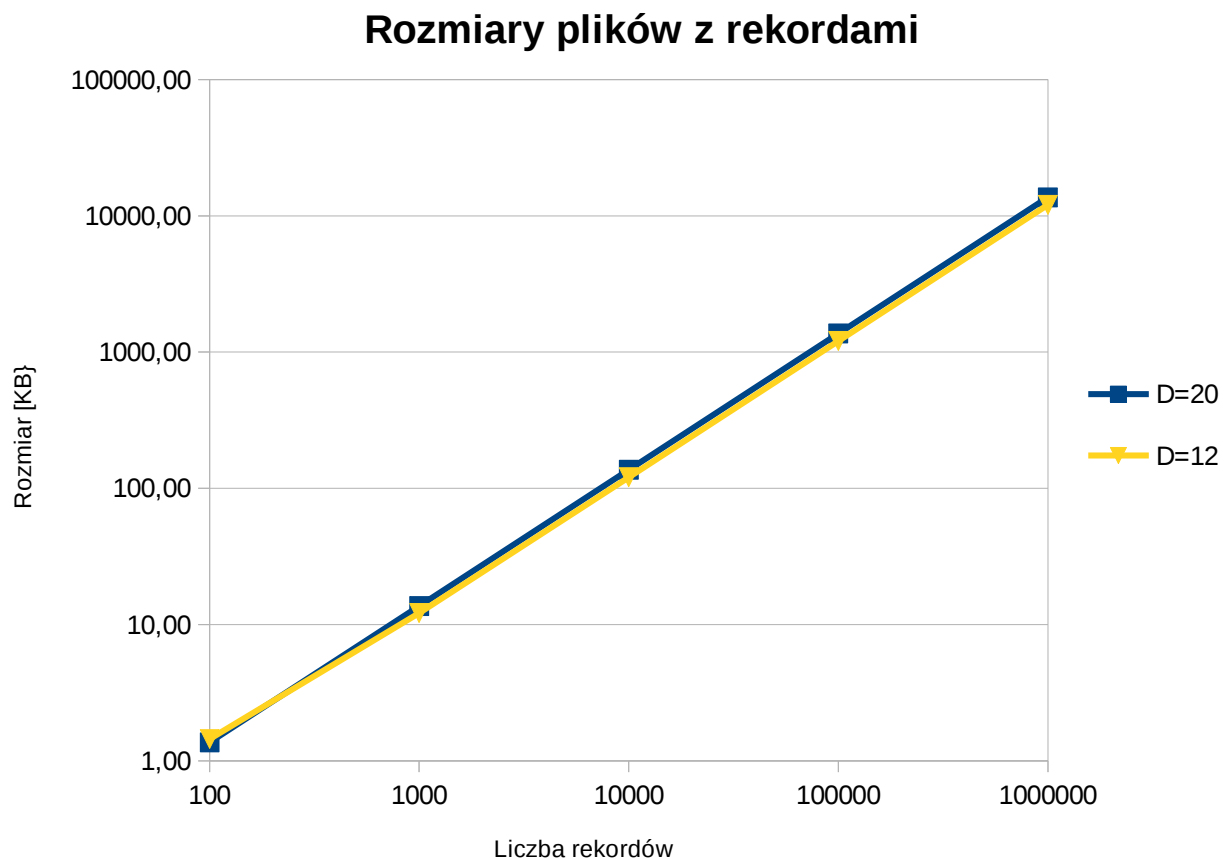
D=12

Liczba rekordów	Index	Records
100	1,73	1,45
1000	13,01	12,14
10000	123,14	120,54
100000	1212,33	1204,52
1000000	12214,05	12044,37

D=20

Liczba rekordów	Index [KB]	Records [KB]
100	1,48	1,37
1000	14,66	13,67
10000	141,37	136,72
100000	1397,65	1367,19
1000000	15096,16	13671,88





Na podstawie wykresów można stwierdzić, że pliki liniowo zwiększają swoje rozmiary, zatem są one prawidłowe. Widać wyraźnie, że rozmiary plików z rekordami nie różnią się bardzo rozmiarami w zależności od rozmiaru stron, czyli od ***d***. Inaczej to wygląda w przypadku indeksów. Zwiększenie rozmiaru strony zmniejsza liczbę stron, a co za tym idzie, mniejsza jest liczba dodanych przeze mnie do każdej strony nagłówków z liczbą kluczy na stronie.