

# STRUKTURY BAZ DANYCH

## Zadanie 1

### Sortowanie pliku metodą sortowania polifazowego

Rekordy pliku: Daty.

#### 1. Opis metody:

Plik wejściowy jest rozdzielany początkowo na dwie taśmy zrealizowane jako pliki według schematu: Dopóki są serie do dodania dodaj na taśmę z mniejszą liczbą serii tyle serii, ile jest aktualnie na taśmie z większą liczbą serii. W przypadku, gdy liczba pozostałych w pliku wejściowym serii okaże się mniejsza niż być powinna, zostaje oznaczona liczba serii pustych tzw. „dummy”. W następnym kroku z krótszej (zawierającej mniej serii) taśmy przepisywane są serie na taśmę wyjściową. Dalej wykonywane jest scalanie polegające na złączeniu wszystkich serii z krótszej taśmy i tyle samo z dłuższej i zapisanie ich na taśmę wyjściową. Taśma krótka staje się tym samym taśmą wyjściową, taśma długa krótką, a wyjściowa długą. Scalanie powtarzane jest do momentu, gdy zostaje po jednej serii na dwóch taśmach. W tym momencie obie serie są scalane i zapisywane w pliku wejściowym. Wszystkie zapisy i odczyty realizowane są blokowo.

#### 2. Format pliku:

W pliku zapisane są rekordy w postaci binarnej, każdy rekord zajmuje 8 bajtów na 1. bajcie zostaje zapisany dzień miesiąca, na trzecim miesiąc, na pozostałych 4 rok, wszystko w formacie U2. Plik zawiera same rekordy, w pliku nie jest zapisywana liczba rekordów.

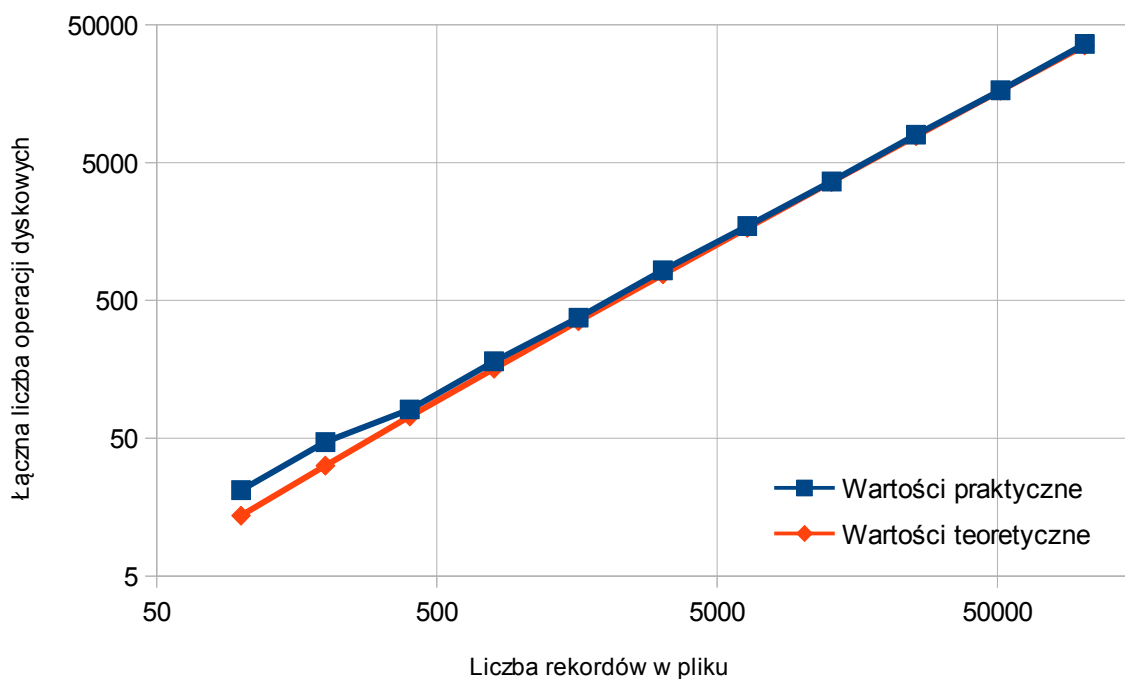
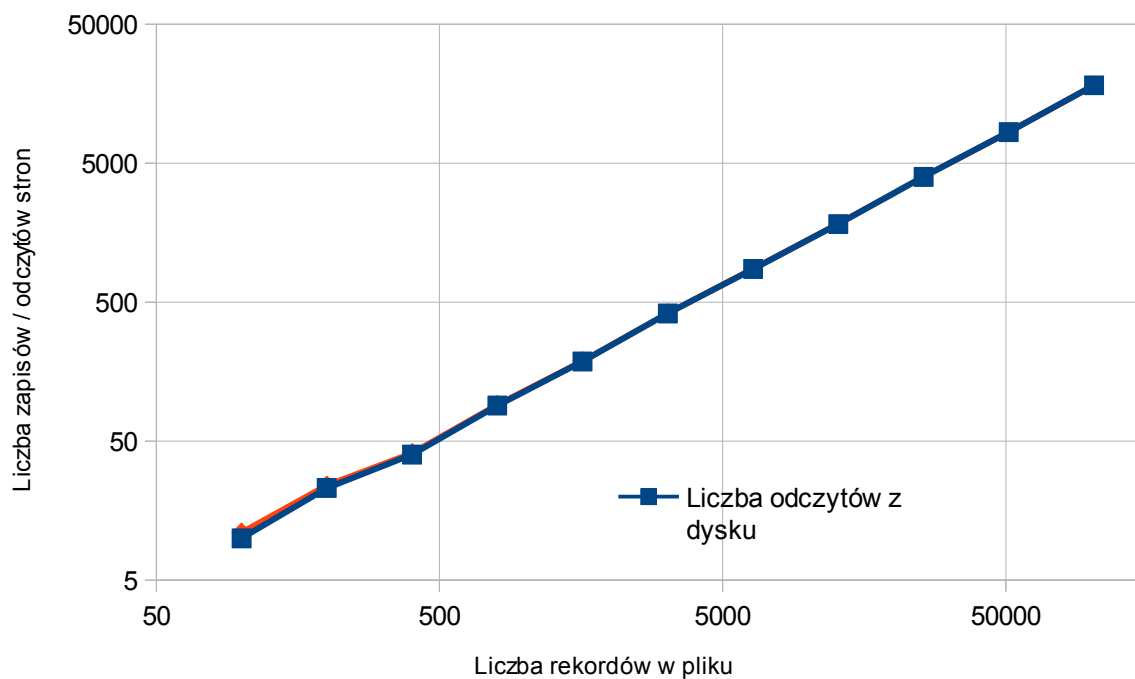
#### 3. Prezentacja wyników działania programu:

Przed rozpoczęciem działania algorytmu zawartość pliku jest wyświetlana wraz z numeracją rekordów i serii. Podczas działania algorytmu, jeśli użytkownik wyraził taką chęć, wyświetlany jest aktualny stan taśm. Po zakończeniu wyświetlany jest posortowany plik jak również liczba wykonanych zapisów i odczytów oraz liczba przeprowadzonych faz algorytmu. Do liczby odczytów nie wliczam odczytów związanych z wyświetlaniem stanu taśm.

#### 4. Eksperyment: Badanie zależności liczby rekordów

W każdej z prób współczynnik blokowania wynosił  $b = 100$ . Program tworzył losowe pliki zawierające od 100 do 102400 rekordów, przy czym z każdą próbą liczba rekordów była podwajana.

Liczba rekordów w pliku	Początkowa liczba serii	Liczba odczytów z dysku	Liczba zapisów na dysk	Łączna liczba operacji dyskowych	Obliczona teoretyczna liczba operacji dyskowych	Liczba faz algorytmu	Obliczona teoretyczna liczba faz algorytmu
100	50	10	11	21	14	8	8
200	100	23	24	47	32	10	10
400	206	40	41	81	72	11	11
800	402	90	91	181	160	13	13
1600	808	187	188	375	353	14	14
3200	1610	414	415	829	773	16	15
6400	3196	865	867	1732	1678	17	17
12800	6395	1823	1824	3647	3622	18	18
25600	12842	3985	3986	7971	7780	20	20
51200	25604	8359	8360	16719	16619	21	21
102400	51185	18140	18141	36281	35367	23	23



## 5. Komentarz

Liczba zapisów jest zawsze niemal równa liczbie odczytów (ok. 1-2 więcej zapisów niż odczytów). Wartości teoretyczne są zawsze niższe niż praktyczne. Wynika to z faktu że nie zawsze odczytywane lub zapisywane całe strony, co również tłumaczy dlaczego ta różnica jest bardziej zauważalna dla mniejszej liczby rekordów.