

# Implementacja i analiza indeksowo-sekwencyjnej organizacji plików

Wojciech Trapkowski

8 grudnia 2024

## 1 Wprowadzenie

Celem projektu jest implementacja i analiza indeksowo-sekwencyjnej organizacji plików (ISAM - Indexed Sequential Access Method), która łączy zalety dostępu sekwencyjnego i bezpośredniego do danych. Metoda ta została opracowana przez IBM w latach 60-tych i do dziś stanowi podstawę wielu systemów bazodanowych.

Podstawowe założenia tej organizacji plików obejmują:

- Przechowywanie danych w uporządkowanej sekwencji rekordów, pogrupowanych w strony o stałym rozmiarze
- Utrzymywanie oddzielnego pliku indeksowego, zawierającego klucze i wskaźniki do odpowiadających im stron w pliku głównym
- Wykorzystanie obszaru przepełnień do obsługi nowych rekordów, których nie można umieścić w pierwotnie przydzielonych stronach
- Okresową reorganizację pliku w celu optymalizacji jego struktury

Organizacja ISAM zapewnia wydajne operacje wyszukiwania dzięki indeksom, zachowując jednocześnie możliwość sekwencyjnego przetwarzania danych. Jest szczególnie efektywna w systemach, gdzie stosunek operacji odczytu do zapisu jest wysoki, a dane są względnie statyczne.

## 2 Struktury plików

- Struktura pliku indeksowego zawiera strony indeksowe, gdzie każdy wpis składa się z klucza początkowego oraz wskaźnika do odpowiedniej strony w pliku głównym.
- Struktura pliku głównego składa się z nagłówka zawierającego liczbę stron oraz wskaźnik do obszaru przepełnień (strażnika), a następnie sekwencji stron zawierających rekordy.
- Obszar przepełnień służy do przechowywania rekordów, które nie mogą być umieszczone w pierwotnie przydzielonych stronach. Każdy rekord w obszarze głównym może wskazywać na dodatkowe rekordy w obszarze przepełnień.
- Organizacja rekordów w stronach opiera się na strukturze Page, która zawiera stałą liczbę wpisów. Każdy wpis zawiera klucz, wartość (PE-SEL), wskaźnik do obszaru przepełnień oraz flagę usunięcia.

## 3 Szczegóły implementacyjne

### 3.1 Buforowanie w pamięci operacyjnej

- Mechanizm buforowania zaimplementowany jest w klasie PageBuffer, która wykorzystuje inteligentne wskaźniki do zarządzania stronami w pamięci.
- Wielkość bufora jest określona przez stałą, która definiuje maksymalną liczbę stron przechowywanych jednocześnie w pamięci.
- Strategia zastępowania stron opiera się na liczbie referencji do strony - usuwane są strony z pojedynczą referencją. Przed usunięciem strony z bufora, jej zawartość jest zapisywana na dysk.
- System śledzi liczbę operacji odczytu i zapisu poprzez liczniki.

### 3.2 Parametry implementacyjne

- Rozmiar strony (PAGE\_SIZE) jest stałą określającą liczbę rekordów w pojedynczej stronie.

- Współczynnik wypełnienia ( $\alpha$ ) określa maksymalną liczbę rekordów w stronie po reorganizacji.
- Współczynnik obszaru przepełnień ( $\beta$ ) definiuje stosunek rozmiaru obszaru przepełnień do obszaru głównego
- Reorganizacja jest wykonywana gdy liczba rekordów w obszarze przepełnień przekroczy ustalony próg ( $\gamma$ ).

## 4 Format pliku testowego

### 4.1 Struktura rekordu

W implementacji rekord jest reprezentowany jako pojedyncza liczba całkowita typu `uint64_t`, przechowująca numer PESEL.

## 5 Prezentacja wyników

### 5.1 Interfejs użytkownika

Program oferuje interaktywny interfejs wiersza poleceń oraz możliwość wykonywania komend z pliku. Dostępne są następujące tryby pracy:

- Tryb interaktywny - oznaczony znakiem zachęty `»`
- Tryb wsadowy - wykonywanie komend z pliku

### 5.2 Dostępne komendy

Program obsługuje następujące polecenia:

- `insert <klucz> <wartość>` - wstawia nowy rekord
- `update <klucz> <wartość>` - aktualizuje istniejący rekord
- `search <klucz>` - wyszukuje rekord o podanym kluczu
- `remove <klucz>` - usuwa rekord o podanym kluczu
- `print` - wyświetla zawartość całej bazy danych

- `print_stats` - wyświetla statystyki (liczba operacji I/O)
- `generate <liczba_kluczy>` - generuje zadaną liczbę losowych rekordów
- `reorganise` - wymusza reorganizację struktury
- `flush` - wymusza zapis buforowanych danych na dysk
- `help` - wyświetla listę dostępnych komend
- `exit/quit` - kończy działanie programu

### 5.3 Format wyświetlania

- Wyniki wyszukiwania są wyświetlane w formacie: wartość rekordu lub komunikat "Not found" dla niezalezionych kluczy
- Błędy operacji są sygnalizowane odpowiednimi komunikatami
- Statystyki pokazują liczbę operacji odczytu i zapisu wykonanych na dysku

## 6 Eksperymenty

### 6.1 Metodologia

- Opis przeprowadzonych testów
- Badane parametry
- Mierzone metryki

### 6.2 Wyniki

#### 6.2.1 Wpływ rozmiaru strony

[Tu umieść wykresy i analizę wpływu rozmiaru strony]

#### 6.2.2 Wpływ współczynnika $\alpha$

[Tu umieść wykresy i analizę wpływu współczynnika wypełnienia]

### 6.2.3 Wpływ współczynnika $\beta$

[Tu umieść wykresy i analizę wpływu rozmiaru obszaru przepelnień]

## 6.3 Analiza złożoności operacji

- Wstawianie
- Wyszukiwanie
- Usuwanie
- Reorganizacja

## 7 Wnioski

- Optymalne wartości parametrów
- Zalecane zastosowania
- Ograniczenia metody

## 8 Bibliografia

### Literatura