

Katedra Inżynierii Oprogramowania  
Wydział Elektroniki, Telekomunikacji i Informatyki  
Politechnika Gdańska

# **STRUKTURY BAZ DANYCH**

PROJEKT

Instrukcja

Opracowali:  
dr hab. inż. Krzysztof Goczyła, prof. PG  
mgr inż. Teresa Grabowska  
mgr inż. Grzegorz Gołaszewski

Gdańsk, 2007

## 1. Cel projektu

Celem projektu jest wyrobienie praktycznej umiejętności w organizowaniu pamięci zewnętrznej bazy danych w struktury plikowe oraz w eksperymentowaniu na tych organizacjach.

## 2. Zadania do realizacji

W ramach projektu studenci wykonują dwa lub trzy zadania, których treść zamieszczono poniżej. Zadania należy zrealizować w języku ogólnego przeznaczenia (Pascal, C, C++ lub Java) i do zaliczenia dostarczyć wersję wykonywalną i źródłową programu oraz przykładowe zbiory danych testowych, zapisane w plikach testowych. We wszystkich zadaniach programowi powinno towarzyszyć pisemne sprawozdanie, które, prócz nazwiska i imienia autora i treści zadania, powinno zawierać informacje specyficzne dla zadania (patrz: treści zadań).

*Uwaga 1:* We wszystkich zadaniach operacje odczytu z dysku i zapisu na dysk należy przeprowadzać stronami (blokowo).

*Uwaga 2:* We wszystkich zadaniach obowiązuje wersja algorytmu przedstawiona na wykładzie.

### Zadanie 1.

Napisz program sortujący plik metodą scalania naturalnego lub metodą sortowania polifazowego. Taśmy powinny być zrealizowane w formie plików dyskowych. Rekordy do posortowania powinny być generowane zarówno losowo, jak i z klawiatury. Program powinien wyświetlić zawartość pliku przed posortowaniem 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.

*Uwaga:* Typ rekordu pliku należy pobrać u prowadzącego projekt.

Sprawozdanie ma zawierać:

- krótki opis zastosowanej metody (w szczególności – liczbę zastosowanych taśm),
- opis eksperymentów,
- wyniki liczbowe eksperymentów dla min. 5 znacznie różniących się od siebie liczb rekordów w pliku  $N$ ,
- wykres łącznej liczby odczytów i łącznej liczby zapisów w funkcji  $N$ ,
- komentarz do wyników uwzględniający w szczególności porównanie wyników praktycznych z teoretycznymi,
- specyfikację formatu pliku testowego,
- opis sposobu prezentacji wyników działania programu.

Termin realizacji zadania: 5 tygodni od daty wydania danych do zadania.

### Zadanie 2.

Zaprojektuj i zaimplementuj wybraną indeksową organizację pliku (B-drzewo lub indeksowo-sekwencyjną). Zrealizuj operacje: wstawiania rekordu, aktualizacji rekordu, odczytu rekordu, usuwania rekordu, przeglądania całej zawartości pliku i indeksu zgodnie z kolejnością wartości klucza oraz operację reorganizacji pliku (jeśli wybrana organizacja tego wymaga). Typ rekordu pliku przyjmij jako typ rekordu z zadania 1 rozszerzony o klucz. Po każdej przeprowadzonej operacji podawaj liczbę faktycznie zrealizowanych operacji odczytu lub zapisu stron dyskowych. Program powinien dawać możliwość wyświetlania zawartości pliku i indeksu po każdej operacji zmieniającej zawartość pliku (tj. po wstawieniu, aktualizacji lub usunięciu rekordu). Dodatkowo program powinien dawać możliwość wczytywania danych testowych z pliku testowego. Danymi testowymi powinna być dowolna sekwencja operacji wstawiania, aktualizacji i usuwania rekordu.

Sprawozdanie powinno zawierać:

- informację, jaką metodę indeksowania przyjęto do implementacji,
- opis implementacji przyjętej metody (zasady buforowania w pamięci operacyjnej, ograniczenia, ew. usprawnienia itp.),
- opis eksperymentów,
- wyniki przeprowadzonych eksperymentów (średnia liczba odczytów i zapisów dyskowych na jedną operację, zajętość pamięci przez plik i indeks w funkcji liczby rekordów i in.),
- komentarz na temat implementacji i uzyskanych wyników,
- specyfikacja formatu pliku testowego,
- opis sposobu prezentacji wyników działania programu.

Termin realizacji zadania: 5 tygodni od daty zakończenia zadania pierwszego.

### Zadanie 3. – dodatkowe

Zaprojektuj i zaimplementuj organizację pliku opartą na rozpraszaniu dynamicznym (rozszerzalnym lub liniowym). Zrealizuj operacje: wstawiania rekordu, aktualizacji rekordu, odczytu rekordu, usuwania rekordu oraz przeglądania całej zawartości pliku. Typ rekordu pliku przyjmij ten sam co w zadaniu 2. Po każdej przeprowadzonej operacji podawaj liczbę faktycznie zrealizowanych operacji odczytu lub zapisu stron dyskowych. Program powinien dawać możliwość wyświetlania zawartości pliku i skorowidza (dla metody rozszerzalnej) po każdej operacji zmieniającej zawartość pliku (tj. po wstawieniu, aktualizacji lub usunięciu rekordu). Dodatkowo program powinien dawać możliwość wczytywania danych testowych z pliku testowego. Danymi testowymi powinna być dowolna sekwencja operacji wstawiania, aktualizacji i usuwania rekordu.

Sprawozdanie powinno zawierać:

- informację, jaką metodę rozpraszania przyjęto do implementacji,
- opis implementacji przyjętej metody (zasady buforowania w pamięci operacyjnej, ograniczenia, usprawnienia itp.),
- opis eksperymentów,
- specyfikację formatu pliku testowego,
- opis sposobu prezentacji wyników działania programu.

Termin oddania opracowania: 2 tygodnie od daty zakończenia zadania drugiego.

### **3. Zasady oceniania**

Warunkiem zaliczenia laboratorium jest zaliczenie (min. 50% punktów) pierwszych dwóch zadań oraz uzyskanie łącznie ponad 50% punktów. Za pierwsze zadanie można otrzymać 30 punktów, w tym 20 punktów za program i 10 punktów za sprawozdanie. Za drugie zadanie można otrzymać 40 punktów, w tym 30 punktów za program i 10 punktów za sprawozdanie. Opóźnienie o każdy tydzień powoduje odjęcie 5 punktów. Zadania 3 nie można oddawać ze spóźnieniem. Przy ocenianiu zadań brana jest pod uwagę jakość przyjętego rozwiązania, liczba i sposób przeprowadzenia eksperymentów oraz treść sprawozdania. Sposób realizacji interfejsu użytkownika ma mniejsze znaczenie. W zadaniu 1 dodatkowo (5 punktami) premiowane jest sortowanie polifazowe. W zadaniu 2 dodatkowo (5 punktami) premiowany jest indeks oparty na B-drzewie. Ponadto można otrzymać 30 punktów za wykonanie zadania trzeciego (łącznie ze sprawozdaniem). 75 punktów stanowi 100%. Maksymalna liczba punktów do zdobycia wynosi 110.