

Katedra Inżynierii Oprogramowania
Wydział Elektroniki, Telekomunikacji i Informatyki
Politechnika Gdańska

STRUKTURY BAZ DANYCH

PROJEKT

Instrukcja

Opracowali:

dr hab. inż. Krzysztof Goczyla, prof. PG
dr inż. Teresa Zawadzka
mgr inż. Grzegorz Gołaszewski

Gdańsk, 2015

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++, 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: We wszystkich zadaniach obowiązuje wersja algorytmu przedstawiona na wykładzie.

Zadanie 1.

Napisz program sortujący plik metodą scalania naturalnego lub scalania z użyciem wielkich buforów, 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),
- specyfikację formatu pliku testowego,
- opis sposobu prezentacji wyników działania programu.

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

Punktacja:

- Poprawna implementacja wybranego algorytmu:
 - **15 punktów** przy implementacji scalania naturalnego,
 - **20 punktów** przy implementacji scalania z użyciem wielkich buforów lub sortowania polifazowego.
- Przygotowanie sprawozdania – **2 punkty**.
- Przeprowadzenie eksperymentu dotyczącego zaimplementowanej metody oraz opisanie go w sprawozdaniu – **8 punktów**,
w opisie tym powinny znaleźć się:
 - opis konstrukcji/sposobu przeprowadzenia eksperymentów,
 - wyniki liczbowe eksperymentów dla min. 5 znacznie różniących się od siebie liczb rekordów w pliku N ,
 - wykres łącznej liczby odczytów rekordów/stron dyskowych i łącznej liczby zapisów rekordów/stron dyskowych w funkcji N ,
 - komentarz do wyników uwzględniający w szczególności porównanie wyników praktycznych z teoretycznymi.
- Zaimplementowanie w algorytmie mechanizmu zapisu i odczytu rekordów stronami – **5 punktów**.

Zadanie 2.

Zaprojektuj i zaimplementuj wybraną indeksową organizację pliku (B-drzewo, B+-drzewo lub indeksowo-sekwencyjną). Zrealizuj operacje: wstawiania rekordu, odczytu rekordu, przeglądania całej zawartości pliku i indeksu zgodnie z kolejnością wartości klucza oraz operację reorganizacji pliku (jeśli wybrana organizacja tego wymaga). Operacja reorganizacji powinna być wywoływana automatycznie w miarę potrzeb oraz powinna istnieć możliwość reorganizacji na żądanie. Typ rekordu pliku przyjmij jako typ rekordu z zadania 1 rozszerzony o klucz (liczbę naturalną). 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 (jeśli zostały zaimplementowane).

Sprawozdanie powinno zawierać:

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

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

Punktacja:

- Poprawna implementacja wybranego algorytmu - **20 punktów**.
- Przygotowanie sprawozdania – **2 punkty**
- Przeprowadzenie eksperymentu dotyczącego zaimplementowanej metody oraz opisanie go w sprawozdaniu – **8 punktów**,
w opisie tym powinny znaleźć się:
 - opis eksperymentów,
 - wyniki przeprowadzonych eksperymentów (średnia liczba odczytów i zapisów stron dyskowych na jedną operację, zajętość pamięci przez plik i indeks w funkcji liczby rekordów i in.),
 - komentarz na temat implementacji i uzyskanych wyników.
- Zaimplementowanie w wybranym algorytmie operacji usuwania i aktualizacji rekordu – **10 punktów**.
- Zaimplementowanie mechanizmu pozwalającego na ponowne wykorzystanie zwalnianego przy usuwaniu miejsca (tylko dla B-drzewa i B+drzewa). – **5 punktów**.

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.),
- specyfikację formatu pliku testowego,
- opis sposobu prezentacji wyników działania programu.

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

Punktacja: **30 punktów** za wykonanie zadania.

3. Zasady oceniania

Warunkiem zaliczenia laboratorium jest zaliczenie pierwszych dwóch zadań. Minimum potrzebne do zaliczenia zadań pierwszego i drugiego to odpowiednio 16 i 22 punkty. Nie jest konieczne wykonanie wszystkich elementów zadania – wystarczy osiągnąć wymaganą minimalną liczbę punktów, z zastrzeżeniem jednak, że punkty podane dla poszczególnych elementów zadania stanowią ocenę maksymalną a faktyczna ocena może, w zależności od jakości wykonania, być niższa.

Oddawanie zadań ze spóźnieniem jest możliwe dla pierwszych dwóch zadań, jednak opóźnienie o każdy rozpoczęty tydzień powoduje odjęcie 5 punktów. Ponadto spóźnienie może wynieść maksymalnie 2 tygodnie. 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.

Przy wyznaczaniu oceny 75 punktów stanowi 100%. Maksymalna liczba punktów do zdobycia wynosi 110.

W razie niezaliczenia któregoś z obowiązkowych zadań możliwe jest jego poprawianie po uprzednim uzgodnieniu z prowadzącym warunków owej poprawy. Poprawiając można zdobyć minimalną liczbę punktów za zadanie (odpowiednio 16 i 22). Poprawiać zadania można do końca semestru (należy się rozliczyć przed rozpoczęciem sesji podstawowej).