

Лабораторна робота № 4. Зовнішнє сортування.

Мета роботи: Навчитись сортувати файли даних, вміст яких перевищує об'єм оперативної пам'яті.

Теоретичні відомості.

Сортування прямим злиттям.

Відсортувати файл А (8, 23, 5, 65, 44, 33, 1, 6), використовуючи зовнішнє сортування прямим злиттям. Наведемо розв'язок у таблиці:

Початковий стан файлу А	8, 23, 5, 65, 44, 33, 1, 6
Крок 1. Розподіл Файл В Файл С Файл А	$\begin{array}{cccc} 8 & 5 & 44 & 1 \\ 23 & 65 & 33 & 6 \\ \hline \underbrace{8 \ 23}_{a'_1} & \underbrace{5 \ 65}_{a'_2} & \underbrace{33 \ 44}_{a'_3} & \underbrace{1 \ 6}_{a'_4} \end{array}$ $(a_1, a_2) \ (a_3, a_4) \ (a_5, a_6) \ (a_7, a_8)$
Крок 2. Розподіл Файл В Файл С Файл А	$\begin{array}{cccc} 8 & 23 & 33 & 44 \\ 5 & 65 & 1 & 6 \\ \hline \underbrace{5 \ 8 \ 23 \ 65}_{a''_1(a'_1, a'_2)} & & \underbrace{1 \ 6 \ 33 \ 44}_{a''_2(a'_3, a'_4)} & \end{array}$
Крок 3. Розподіл Файл В Файл С Файл А	$\begin{array}{l} 5 \ 8 \ 23 \ 65 \ (a''_1) \\ 1 \ 6 \ 33 \ 44 \ (a''_2) \\ 1 \ 5 \ 68 \ 23 \ 33 \ 44 \ 65 \end{array}$

Для виконання зовнішнього сортування методом прямого злиття в основній пам'яті необхідно розташувати лише дві змінні – для розміщення записів з файлів В і С.

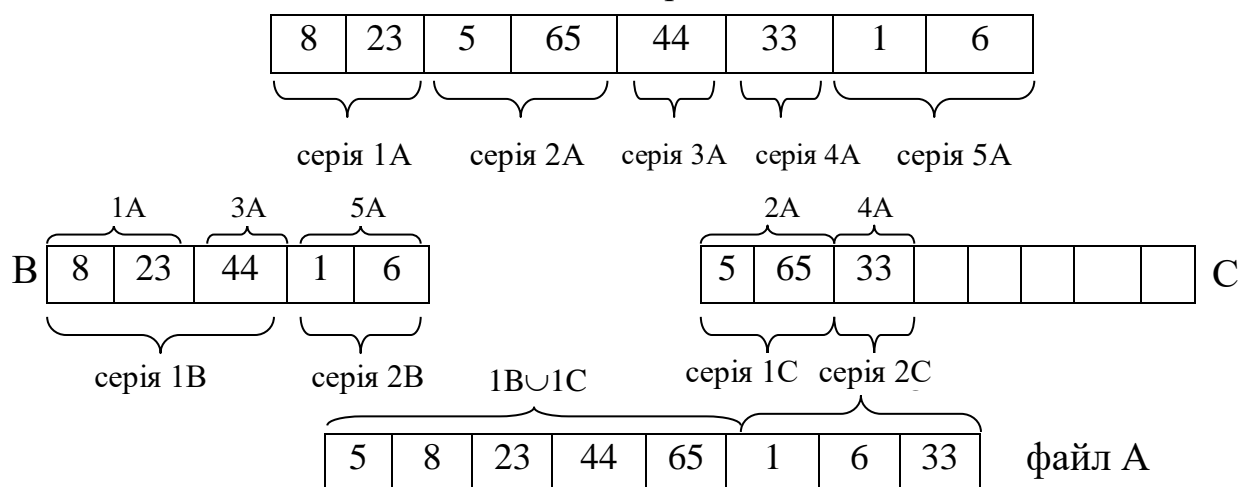
Природне злиття.

Сортування виконується за декілька кроків, в кожному з яких спочатку виконується розподілення файлу А по файлах В і С, а потім

злиття В і С в файл А. При розподілі розпізнається перша серія записів і переписується в файл В, друга – в файл С і т. д. При злитті перша серія записів файлу В зливається з першою серією файлу С, друга серія В з другою серією С і т. д. Якщо перегляд одного файлу закінчується раніше ніж перегляд другого (з причини різної кількості серій), то залишок недопереглянутого файлу ціяко переноситься в кінець файлу А. Процес завершується коли в файлі А залишається лише одна серія (рис. 15).

Оскільки довжина серій може бути довільною, то максимальний розмір файлів В і С може бути близьким до розміру файлу А.

Крок 1.



Крок 2.

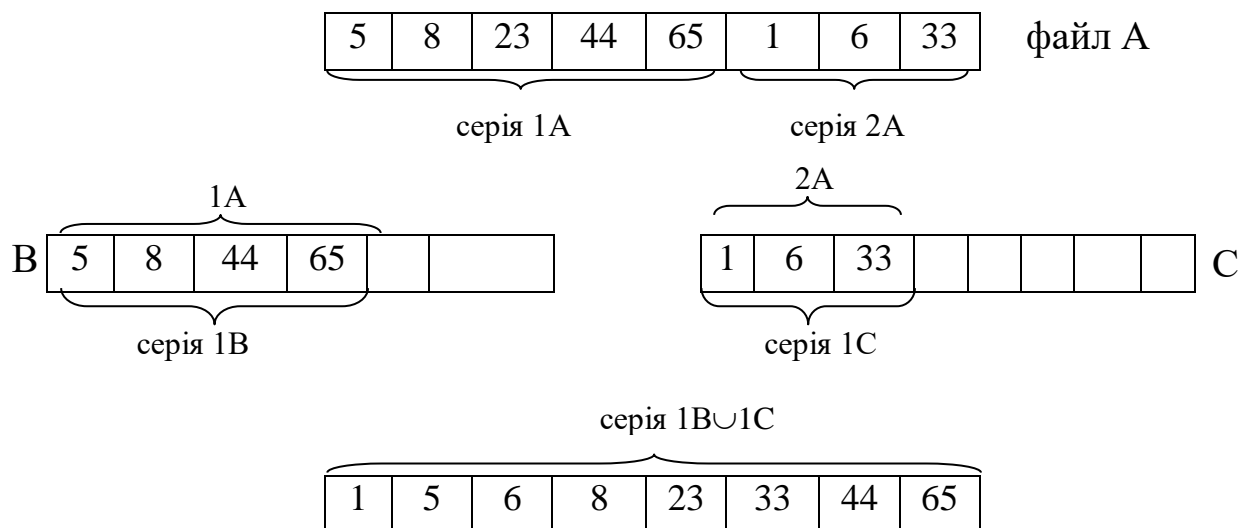


Рис. 15. Природне злиття

Збалансоване багатоканальне злиття.

В основі методу зовнішнього сортування збалансованим багатоканальним злиттям є розподіл серій вхідного файлу по m допоміжних файлів C_1, C_2, \dots, C_m і т. д., доки в B_1 або C_1 не утвориться одна серія. Відсортувати файл, використовуючи трьохканальне злиття (рис. 16).

Як видно з рис. 16, по мірі, збільшення довжини серій допоміжні файли, з великими номерами (починаючи з номера n) перестають використовуватися, оскільки їм „не дістається” жодної серії.

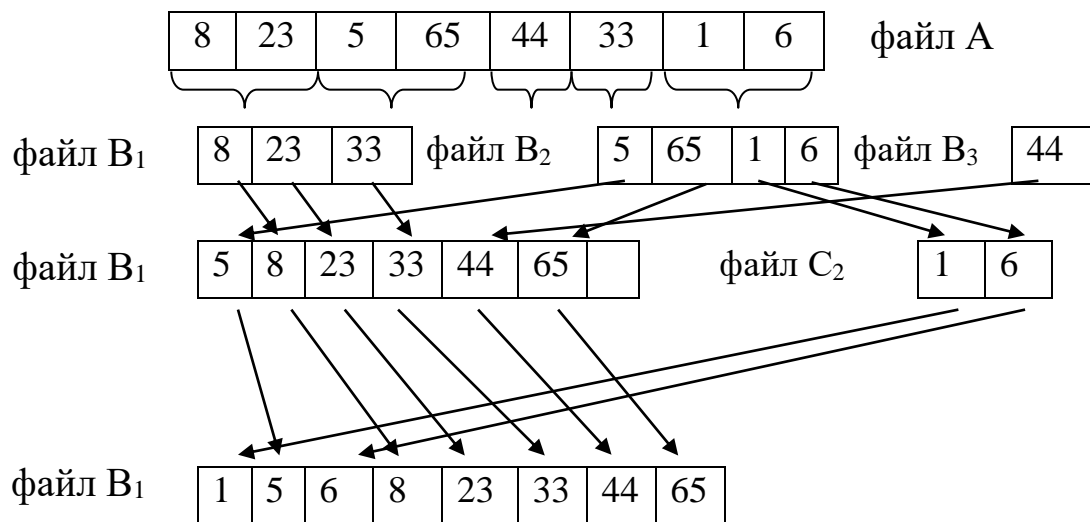


Рис. 16. Трьохканальне злиття

Багатофазне сортування.

Ідея багатофазного сортування полягає в тому, що з m допоміжних файлів $(m-1)$ файл використовується для вводу послідовностей, що зливаються, а один для виводу утворених серій. Як тільки один з файлів вводу стає порожнім, його починають використовувати для виводу серій, отриманих при злитті серій нового набору $(m-1)$ файлів. Таким чином, маємо перший крок, при якому серії початкового файлу А розподіляються по $m-1$ допоміжному файлу, а потім виконується багатофазне злиття серій з $(m-1)$ файлів, доки в одному з них не утвориться одна серія.

Відсортувати файл, використовуючи трьохфазне злиття.

Крок	B_1	B_2	B_3
n	1	0	0
$n-1$	0	1	1
$n-2$	1	2	0
$n-3$	3	0	2
$n-4$	0	3	5
$n-5$	5	8	0
$n-6$	13	0	8
...
1

Цей приклад показує, що метод трьохфазного зовнішнього сортування дає бажаний результат і працює максимально ефективно (на кожному етапі відбувається злиття максимальної кількості серій), якщо початковий розподіл серій між допоміжними файлами описується сусідніми числами Фібоначчі.

Завдання.

Скласти схеми алгоритмів і написати програми реалізації алгоритмів зовнішнього сортування (по варіантах):

- а) природне злиття;
- б) пряме злиття;
- в) багатоканальна злиття;
- г) багатofазне злиття.

Варіант	Текст Т
1.	T = afdfabcabcbcfidca
2.	T = cbafcbcbacbcbedcak
3.	T = hdacbacbacabadhac
4.	T = fdfacbacbacbadafc
5.	T = ftfabababadbadfdkt
6.	T = bnbfdbdadbaddnfdn
7.	T = batfkatktkatkkfabk
8.	T = baxazzazzxzzaxaxb
9.	T = npxamnmmpnmmpmnapn
10.	T = nbnamanamnmanabanb
11.	T = batfxyxzyxzxzfabki
12.	T = tpotpptpopptpoapkp
13.	T = ankakbbykbnnakapkp
14.	T = kerpedecedecerkep
15.	T = kacvcmpmcmmpcevck
16.	T = ytrtqqptqqptqytyq
17.	T = ssfaffcfccfcqasfs
18.	T = scbabkfakfakakfbass
19.	T = nbfbegfgegfefenbnf
20.	T = bgagcdcdcdcedegba
21.	T = cvcxmcmcpmcmcpxcz
22.	T = bgnbxzyxzyxzbncz
23.	T = nsabaappkaappkbnas
24.	T = fdghmamapmamaahfgf
25.	T = fcbaslmnlmnlmgbac
26.	T = nvbvmpkmpkpmknbvc
27.	T = tbtdeffcfccfpctbtd
28.	T = scaskfckfckfbasb
29.	T = bnfssrrtrsrrtnbdf
30.	T = baacfkpfkpfkfbdac

Контрольні питання.

1. Де використовується зовнішнє сортування?
2. Від чого залежить швидкість виконання зовнішнього сортування?
3. В чому полягає ідея прямого, природного, багатоканального та багатофазного злиття?
4. Яким повинен бути початковий розподіл серій між допоміжними файлами у багатофазному злитті, щоб алгоритм працював найефективніше?