**Сортировка файлов (внутренняя и внешняя)**

Сортировка – это процесс перегруппировки заданного множества объектов в некотором определённом порядке.

Цель сортировки – облегчить последующий поиск элементов в таком отсортированном массиве (файле).

Методы сортировки разбиты на 2 больших класса:

1. Сортировка массивов (внутренняя сортировка);
2. Сортировка файлов (внешняя сортировка).

Элемент, по которому производится сортировка, называется ключом (элемент массива, множества, файла и т.д.)

**Методы сортировки массивов**

Основное условие – выбираемый метод сортировки массивов должен экономно использовать доступную память. Самые простые методы сортировки – это прямые методы:

1. Прямое включение;
2. Прямой выбор;
3. Прямой обмен («пузырёк»).

**Прямое включение**

В общем виде: элементы мысленно делятся на исходную последовательность и уже отсортированную («готовую») последовательность: a1, … , an. При каждом шаге, начиная с i=2 и увеличивая i каждый раз на 1, из исходной последовательности извлекается i-ый элемент и перекладывается в готовую последовательность, и при этом он вставляется на нужное место.

Procedure sort1 (n: integer);

Var I, j, x: integer;

A: array [0.. n] of integer;

Begin

For i:=2 to n do begin

X:= a[i]; a[0]:= x; j:= I;

While (x < a[j-1]) do begin

A[j]:= a[j-1]; j:= j-1; end;

A[j]:= x; end;

End;

Этот алгоритм можно улучшить.

**Прямой выбор**

1. Выбирается элемент с наименьшим ключом (min);
2. Он меняется местами с первым элементом a1;
3. Затем этот процесс повторяется с оставшимися (n-1) элементами, (n-2) элементами и т.д. до тех пор, пока не останется один, самый большой элемент.

Прямой выбор противоположен прямому включению:

1. При прямом включении на каждом шаге рассматривается только один очередной элемент исходной последовательности и все элементы готовой последовательности, среди которых отыскивается точка включения;
2. При прямом выборе для поиска одного элемента с наименьшим ключом просматриваются все элементы исходной последовательности, и найденный элемент помещается как очередной элемент в готовую последовательность.

Procedure sort2 (n: integer);

Var I, j, k, x: integer;

A: array [1.. n] of integer;

Begin

For i:= 1 to n-1 do begin

K:= I; x:= a[i];

For j:= i+1 to n do begin

If (a[j] < x) then begin

K:= j; x:= a[k]; end; end;

A[k]:= a[i]; a[i]:= x; end;

End;

**Прямой обмен («пузырёк»)**

1. Одна исходная последовательность;
2. Сравнение и обмен местами двух соседних элементов до тех пор, пока не будут упорядочены все элементы. То есть получается, что в массиве наименьший элемент пошагово сдвигается влево. Если массив рассматривать вертикально, то передвижение будет снизу вверх, как будто всплывает пузырёк воды.

Усовершенствованный пузырёк – это когда чередуются направления последовательных просмотров. Он называется шейкерной сортировкой.

**«Пузырёк»**

Procedure sort3 (n: integer);

Var I, j, x: integer;

A: array [1.. n] of integer;

Begin

For i:= 2 to n do begin

For j:=n down to I do begin

If (a[j-1] > a[j]) then begin

X:= a[j-1]; a[j-1]:= a[j]; a[j]:= x;

End; end; end;

End;

**Шейкерная сортировка**

Procedure sort4 (n: integer);

Var j, k, l, r, x: integer;

A: array [1.. n] of integer;

Begin

L:=2; r:=n; k:=n;

Repeat

For j:= r down to l do begin

If (a[j-1] > a[j]) then begin

X:= a[j-1]; a[j-1]:= a[j]; a[j]:= x; k:=j; end; end;

L:= k+1;

For j:= l to r do begin

If (a[j-1] > a[j]) then begin

X:= a[j-1]; a[j-1]:= a[j]; a[j]:= x; k:=j; end; end;

R:= k-1;

Until (l>r);

End;

**Улучшенные методы сортировки:**

1. Сортировка с помощью включений с уменьшающимися расстояниями между элементами;
2. Сортировка с помощью бинарного дерева;
3. Сортировка с помощью разделения массива на части – быстрая сортировка;
4. Нахождение медианы, т.е. элемента, меньшего или равного половине из n элементов и большего или равного другой половине из n элементов.

**Сортировка файлов**

1. Прямое слияние.
2. Естественное слияние.
3. Сбалансированное многопутевое слияние.
4. Многофазная сортировка.
5. Распределение начальных серий.

**Прямое слияние**

Слияние – это объединение двух и более файлов в один упорядоченный файл с помощью повторяющегося выбора из доступных в данный момент элементов.

**Простое слияние**

1. Последовательность A разбивается на две половины: B и C.
2. B и C сливаются так, что одиночные элементы (один из B, другой из C) образуют упорядоченные пары.
3. Полученная последовательность A вновь обрабатывается так же, как и в 1) и 2); при этом упорядоченные пары переходят в упорядоченные четвёрки.
4. Повторяя 1), 2), 3) сливаем четверки в восьмёрки и т.д., каждый раз удваивая длину последовательностей до тех пор, пока не будет упорядочена целиком вся последовательность A.

Действия по однократной обработке всего множества данных называются фазой, а наименьший подпроцесс, повторение которого составляет процесс сортировки, - проходом.

**Естественное слияние**

В случае прямого слияния мы не получаем никакого преимущества, если данные в начале уже частично упорядочены. Размер сливаемых на K-том проходе подпоследовательностей меньше или равен два в степени K, т.е. любые две упорядоченные подпоследовательности длиной m и n можно сразу сливать в одну последовательность из (m+n) элементов. Сортировка, при которой всегда сливаются две самые длинные из возможных подпоследовательностей, называется естественным слиянием.

**Сбалансированное многопутевое слияние**

Если последовательность разбивать не на две подпоследовательности, а на R и сливать уже по R последовательностей, то мы получим сбалансированное многопутевое слияние.

**Многофазная сортировка**

Здесь уже отказываются от проходов и переходят к более изощрённому использованию последовательностей.

**Распределение начальных серий** – это сочетание методов сортировки массивов и файлов.