Внешняя сортировка

- **Принято называть "внешней" сортировкой сортировку последовательных файлов, располагающихся во внешней памяти и слишком больших,**

чтобы можно было целиком переместить их в основную память и применить один из методов внутренней сортировки.

Внешняя сортировка:прямое слияние

- **Имеется последовательный файл A, состоящий из записей a1, a2, ..., an**
- **Для сортировки используются два вспомогательных файла B и C (размер каждого из них - n/2).**
- **Сортировка состоит из последовательности шагов, в каждом из которых выполняется распределение состояния файла A в файлы B и C, а затем слияние файлов B и C в файл A.**

Внешняя сортировка:прямое слияние

- **Шаг1. Последовательно читается файл A, и записи a1, a3, ..., a(n-1) пишутся в файл B, а записи a2, a4, ..., an - в файл C (начальное распределение)**
- **Шаг2. Начальное слияние производится над парами (a1, a2), (a3, a4), ..., (a(n-1), an),**
- **Шаг3. Последовательно читается файл A, и в файл B записываются последовательные пары с нечетными номерами, а в файл C - с четными. (пример)**
- **Шаг4. Слияние B и C образуются и пишутся в файл A упорядоченные четверки записей...**

Внешняя сортировка:естественное слияние

- **Серией называется подпоследовательность записей**

$$a_i, a_{(i+1)}, \ldots, a_j$$

такая, что

$$a_k <= a_{(k+1)} \text{ для всех } i <= k < j,$$

$$a_i < a_{(i-1)}$$

$$\text{и } a_j > a_{(j+1)}.$$

Метод естественного слияния основывается на распознавании серий при распределении и их использовании при последующем слиянии.

Внешняя сортировка:естественное слияние

- **При распределении распознается первая серия записей и переписывается в файл B, вторая - в файл C и т.д.**
- **При слиянии первая серия записей файла B сливается с первой серией файла C, вторая серия B со второй серией C и т.д.**
- **Если просмотр одного файла заканчивается раньше, чем просмотр другого (по причине разного числа серий), то остаток недопросмотренного файла целиком копируется в конец файла A.**
- **Процесс завершается, когда в файле A остается только одна серия. (демо)**

Сбалансированное многопутевое слияние

- **распределение серий исходного файла по m вспомогательным файлам**

B1, B2, ..., Bm

и их слияние в m вспомогательных файлов

C1, C2, ..., Cm.

- **на следующем шаге производится слияние файлов C1, C2, ..., Cm в файлы B1, B2, ..., Bm и т.д., пока в B1 или C1 не образуется одна серия (демо)**

Многофазная сортировка

- **Из имеющихся m вспомогательных файлов (m-1) файл служит для ввода сливаемых последовательностей, а один – для вывода образуемых серий.**
- **Как только один из файлов ввода становится пустым, его начинают использовать для вывода серий, получаемых при слиянии серий нового набора (m-1) файлов.**
- **Таким образом, имеется первый шаг, при котором серии исходного файла распределяются по m-1 вспомогательному файлу, а затем выполняется многопутевое слияние серий из (m-1) файла, пока в одном из них не образуется одна серия.**