

Проект 2. Сортировки

Идея удобного расположения (сортировки) объектов очень древняя. Люди издавна упорядочивали числа, идея упорядочивания лежит в основе изобретения самих чисел как средства пересчёта. Идея расположения слов в определённом порядке была известна уже в Древней Греции. Археологи много раз находили упорядоченные по алфавиту (по одной первой или даже по первым двум буквам) списки жителей древнегреческих поселений. Словарный порядок в современном понимании был впервые описан в 1286 г. Однако и спустя 300 лет этот метод не был повсеместно распространён и требовал дополнительных пояснений.

Раньше при составлении словарей расстановка словарных статей в определённом порядке производилась вручную, поэтому часто допускались ошибки. Впервые сортировка данных при помощи компьютера была реализована в 1946 г. Сегодня уже большинство задач сортировки объектов решается при помощи компьютера. При этом компьютерная программа сортировки может опираться на разные способы сортировки. Именно их мы и будем обсуждать в ходе проекта. Выбор способа сортировки зависит от того, какие именно объекты и в каком именно порядке требуется рассортировать. А главное — как много этих объектов.

Обсуждение способов сортировки, конечно, тебе пригодится не только при составлении программы для компьютера. Ясно, что не любую сортировку человек может поручить компьютеру. Например, сортировку вещей в ящиках стола и сортировку предстоящих дел по степени их важности человек должен выполнять самостоятельно.

Каждый вид сортировки, с которым ты познакомишься в этом проекте, — это последовательность простых действий с объектами сортировки, таких, как:

- сравнить два объекта;
- поменять местами два объекта в списке;
- поставить объект в определённое место общего списка;
- вставить объект между двумя соседними объектами списка и др.

Похожие действия производит и компьютер, когда выполняет программу сортировки. Выполняя такие действия в опре-

делённом порядке, сортировщик (тот, кто выполняет сортировку, будь то человек или машина) рано или поздно получает результат — последовательность объектов, рассортированных нужным образом.

В этом проекте мы будем сортировать числа и слова (в прямом или обратном словарном порядке). Для простоты описание способов сортировки мы даём на примере сортировки чисел в порядке возрастания.

1. Сортировка с помощью прямого выбора

Этот метод состоит в том, чтобы строить последовательность чисел по порядку: первое, второе, третье и т. д. Берём несколько карточек с числами. Выбираем из всех чисел самое маленькое и ставим его в начало нашей последовательности (стопки карточек). Для этого нам придётся прямо или косвенно сравнить это число со всеми остальными числами. После того как первое число найдено, опять остаётся кучка карточек с числами, среди которых мы выбираем наименьшее. Оно будет вторым в нашей последовательности. Так мы будем продолжать, пока все карточки не закончатся.

Такой метод сортировки годится только тогда, когда карточек с числами не слишком много (иначе поиск наименьшего числа каждый раз будет очень долгим) и все числа доступны для просмотра практически одновременно. Если же карточки с числами собраны в стопку и разрешается их перебирать в этой последовательности только по одной, то ясно, что такой метод сортировки не годится (он займёт слишком много времени).

2. Сортировка с помощью прямой вставки

В отличие от сортировки с помощью выбора, здесь мы не ищем каждый раз наименьшее число, а берём первую попавшуюся карточку из кучки и находим для неё правильное место в нашей последовательности.

Итак, берём любую карточку и выкладываем её на стол — начинаем строить последовательность. Затем берём другую карточку (не выбирая из кучки) и ставим её после первой или перед ней так, чтобы порядок был правильный. И дальше поступаем так же: берём какую-нибудь карточку, находим для неё правильное место в нашей последовательности и вставляем

на это место. Так мы делаем, пока карточки с числами не закончатся.

Такой метод сортировки тоже годится только тогда, когда чисел не слишком много (иначе поиск нужного места для вставки числа займёт слишком много времени). Но зато при таком методе неважно, как именно собраны числа вначале. Например, они могут быть собраны в стопку так, что видна только верхняя карточка. Или это может быть непрозрачный мешок с карточками, которые мы достаём по одной не глядя.

3. Сортировка методом «всплывающего пузырька»

В отличие от двух предыдущих методов, здесь мы сразу работаем со всеми карточками, сложенными в стопку. При этом оказывается, что, каждый раз меняя местами лишь две соседних карточки, можно превратить неупорядоченную последовательность в упорядоченную.

Итак, сначала собираем все карточки в стопку. Начинаем изучать нашу стопку карточек снизу, с конца. Смотрим на последнее и предпоследнее числа в стопке. Если эти две карточки лежат в правильном порядке (то число, которое меньше, идёт раньше), то оставляем карточки на месте, если нет — меняем карточки местами. Затем аналогично сравниваем второе и третье числа с конца, дальше — третье и четвертое и т. д., до начала цепочки. После того как мы пришли к началу цепочки, мы снова возвращаемся к первому и второму числам с конца и проходим по всей цепочке ещё раз. Так мы делаем до тех пор, пока все числа не окажутся упорядоченными.

Можно сказать, что в результате каждого прохода более «лёгкие» элементы (т. е. меньшие числа) всплывают, как пузырьки воздуха в воде, поэтому такой метод называется методом «всплывающего пузырька». При такой сортировке «лёгкий» (наименьший) элемент «всплывает» за один проход, а «тяжёлые» (большие) элементы опускаются при каждом проходе только на одну позицию. Поэтому, если в исходной последовательности наибольший элемент находится в самом начале, то потребуется столько проходов, сколько членов в этой последовательности. Из-за этого такой метод годится только для упорядочивания не слишком длинных последовательностей.

Описанные три метода сортировки называются *прямыми*. Они считаются не очень быстрыми, и при их помощи можно упорядочивать только небольшие количества объектов. При использовании прямых методов сортировки не получается разделить работу между несколькими участниками.

Для сортировки очень больших количеств объектов (например, миллиона чисел) используются более быстрые методы сортировки. Один из таких методов так и называется: «метод быстрой сортировки». Если для сортировки на компьютере 1 000 000 чисел методом прямой вставки требуется почти 1 ч работы, то «методом быстрой сортировки» эта же задача на том же компьютере решается меньше, чем за 3 мин! Но быстрые методы сортировки достаточно сложно устроены, их трудно описать простыми словами, ещё труднее объяснить, почему они работают быстрее. Поэтому мы ими заниматься сейчас не будем.

Другой путь сделать сортировку более быстрой — *разделить работу между несколькими участниками* так, чтобы они работали одновременно. Современные компьютеры могут выполнять много задач одновременно, поэтому такие методы сортировки используются и при создании программ для компьютера.

4. Сортировка слиянием

Ключевая идея этого метода состоит в том, что не обязательно сразу строить общую большую стопку карточек — можно вначале построить несколько небольших упорядоченных стопок, а затем соединить их в одну. Сначала делим все карточки на две стопки, примерно одинаковые по числу карточек, потом выдаём эти стопки двум сортировщикам, и они работают одновременно, упорядочивая числа каждый в своей стопке. Полученные упорядоченные стопки кладём на стол так, чтобы в каждой стопке сверху лежала карточка с наименьшим числом. Теперь сливаем две полученные последовательности по следующим правилам. Самое первое число общей стопки будем искать среди первых чисел двух наших маленьких стопок (мы специально положили стопки так, чтобы видеть только наименьшие числа). Берём ту из двух верхних карточек, на которой число меньше, и кладём её в новую (общую) стопку числом вниз. Теперь мы снова видим только два числа — верхние карточки наших двух маленьких стопок — и выбираем второе число об-

щей стопки из этих двух чисел. Так мы будем поступать, пока не закончатся все числа в стопках.

Мы описали, как можно при помощи сортировки слиянием распределить работу между двумя сортировщиками. Конечно, аналогично можно распределить работу и между большим числом сортировщиков: их может быть 4, 8, 16 и т. д. (удобно, когда число получаемых последовательностей на каждом этапе делится на 2). С одной стороны, чем больше сортировщиков работает одновременно, тем меньше чисел достанется каждому на втором этапе и тем быстрее пройдет общая сортировка. С другой стороны, чем больше сортировщиков участвует, тем больше потребуются этапов попарного слияния. Поэтому неясно, что будет быстрее: разбить стопку на 16 частей или только на 4 части и потом собрать общую последовательность слиянием. Это вы сможете выяснить экспериментально.

5. Сортировка через группировку

Такая сортировка состоит в том, чтобы сразу последовательно разделить все карточки на стопки так, чтобы потом слияние не понадобилось: например, в одну стопку сразу откладываем все числа из первой сотни (от 1 до 99), во вторую — все числа из второй сотни (от 100 до 199), в третью — из третьей сотни и т. д. Каждую стопку выдаём одному сортировщику. Теперь все сортировщики работают одновременно и каждый сортирует свою стопку карточек каким-либо прямым методом сортировки. По завершении этого этапа получаем несколько упорядоченных последовательностей чисел, которые осталось только сложить в одну последовательность, поставив стопки одну за другой.

Быстрота сортировки через группировку очень сильно зависит от того, какие именно числа есть в наборе. Вполне может оказаться, что в заданном наборе наибольшее число 1000, но только два числа из всего набора меньше 500, а остальные числа больше 750 (например, в набор могут войти просто все числа от 751 до 1000). Ясно, что в такой ситуации предложенный нами способ группировки не даст никакой экономии времени: работать будет один сортировщик, а трое остальных будут бездельничать. Поэтому этот метод хорош только в тех случаях, когда заранее известен примерный состав набора чисел, который требуется рассортировать.