

Сравнение производительности сортировок на массивах и списках

Зарипов Данил Алмазович

СПбГУ, 2024

Асимптотический анализ

- Асимптотика алгоритма — математическая граница его производительности на основе размера входных данных.
- О-нотация, O («О» большое): описывает верхнюю границу времени (время выполнения «не более, чем...»).
- Алгоритмы могут иметь одинаковую сложность, но при этом значительно отличаться в производительности.

Сортировка

- Алгоритм сортировки – алгоритм для упорядочивания элементов в списке.
- Алгоритм сортировки – классический объект для различного вида оценок времени и эффективности использования памяти.

Постановка задачи

Цель – проанализировать производительность сортировок.

Задачи:

- Выявить самую быструю сортировку среди быстрых($O(n \log n)$) и среди медленных($O(n^2)$).
- Выяснить разницу в производительности аналогичных сортировок на массивах и односвязных списках.

Объекты исследования

Быстрые сортировки:

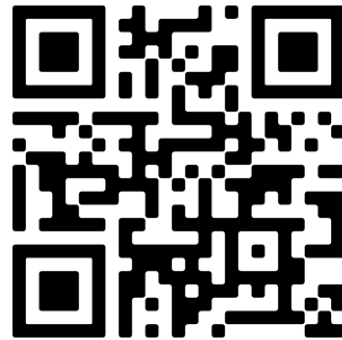
- Сортировка слиянием – худшее, лучшее и среднее время $O(n \log n)$.
- «Быстрая» сортировка – худшее время $O(n^2)$, лучшее и среднее $O(n \log n)$.

Медленные сортировки:

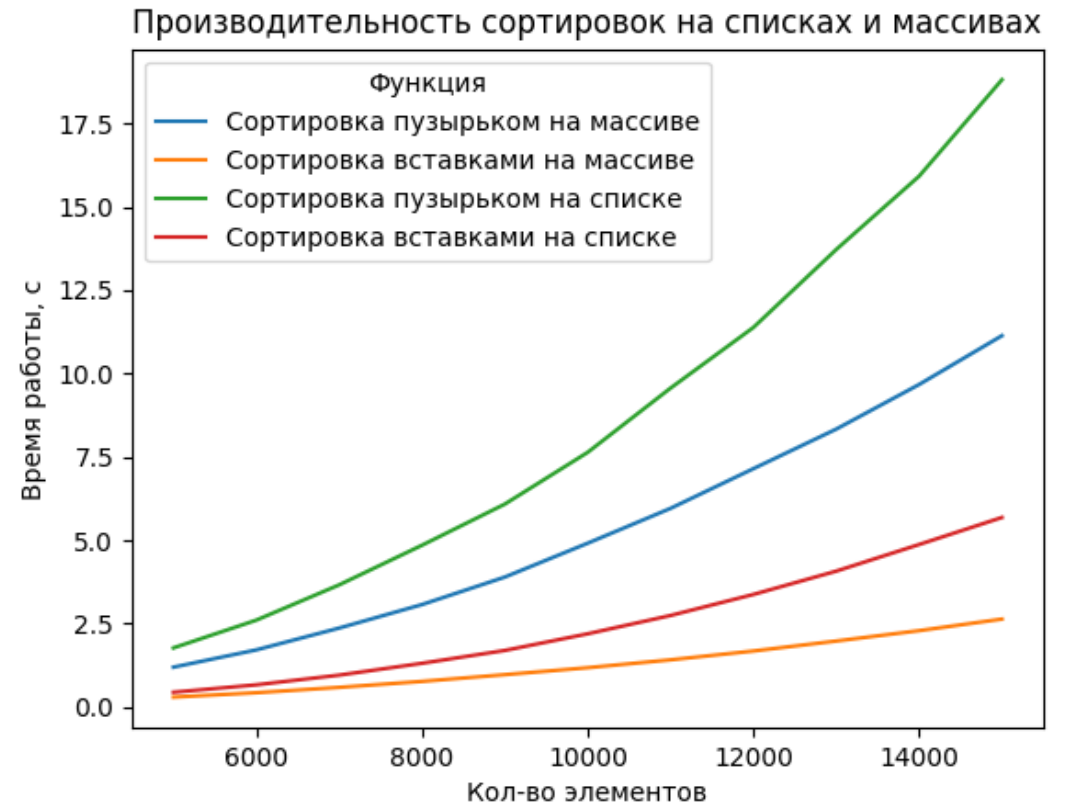
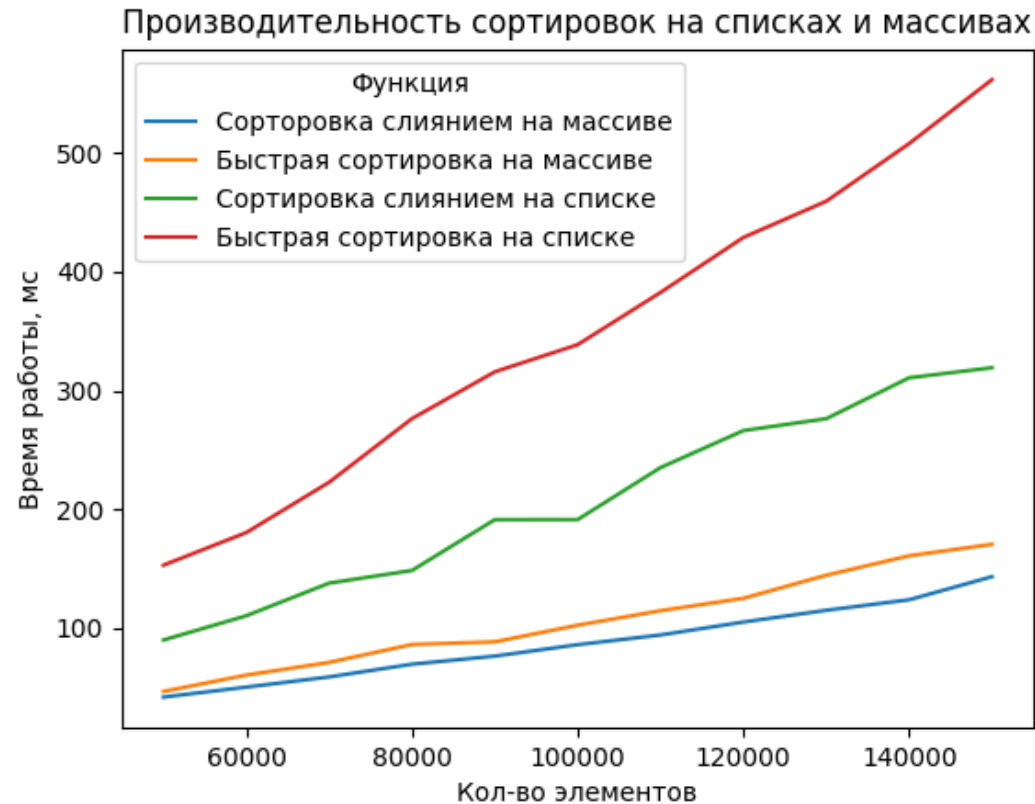
- Сортировка «пузырьком» – худшее, лучшее и среднее время $O(n^2)$.
- Сортировка вставками – худшее, лучшее и среднее время $O(n^2)$.

Как проводился эксперимент

- Платформа .NET, язык F#, библиотека BenchmarkDotNet.
- Для медленных сортировок количество элементов от 5000 до 15000 с шагом в 1000, для быстрых – от 50 тыс. до 150 тыс. с шагом 10000. Элементы генерируются псевдо-случайно.
- Исходный код сортировок и бенчмарков в репозитории:



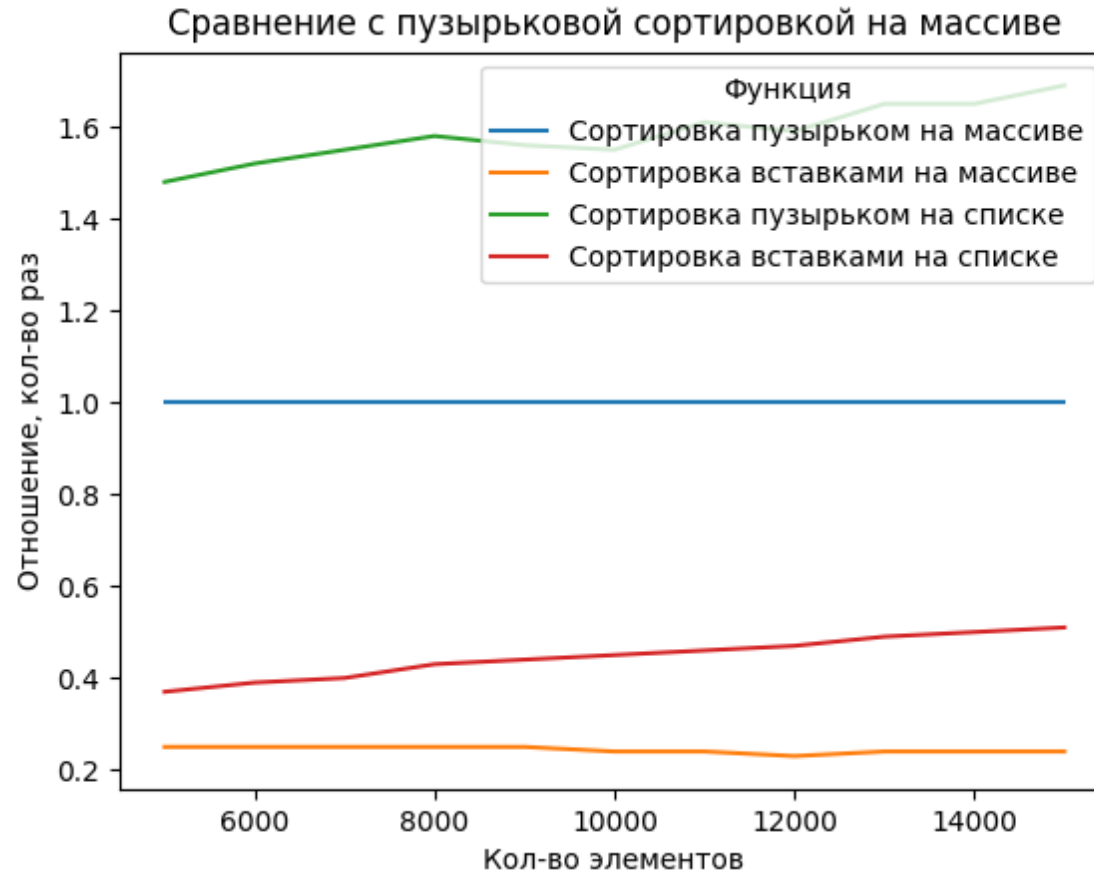
Время работы сортировок



Ниже – лучше.

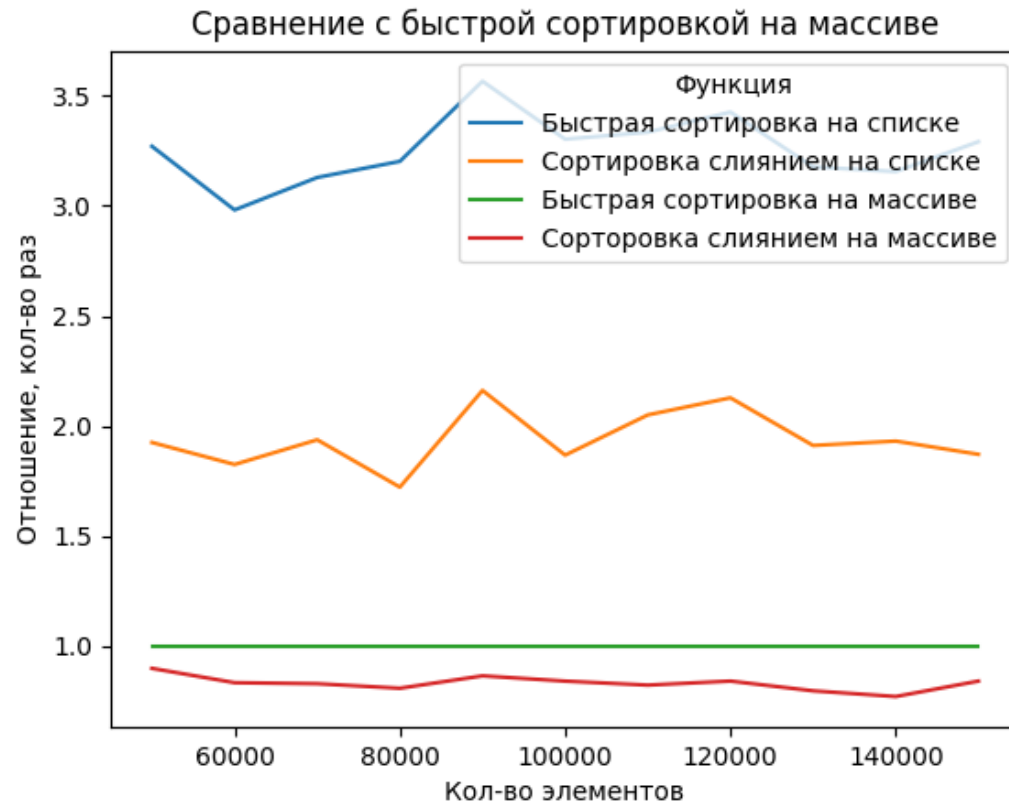
Объем данных для медленных в десять раз меньше, но время работы в сто раз больше.

Сравнение медленных сортировок с сортировкой пузырьком на массиве



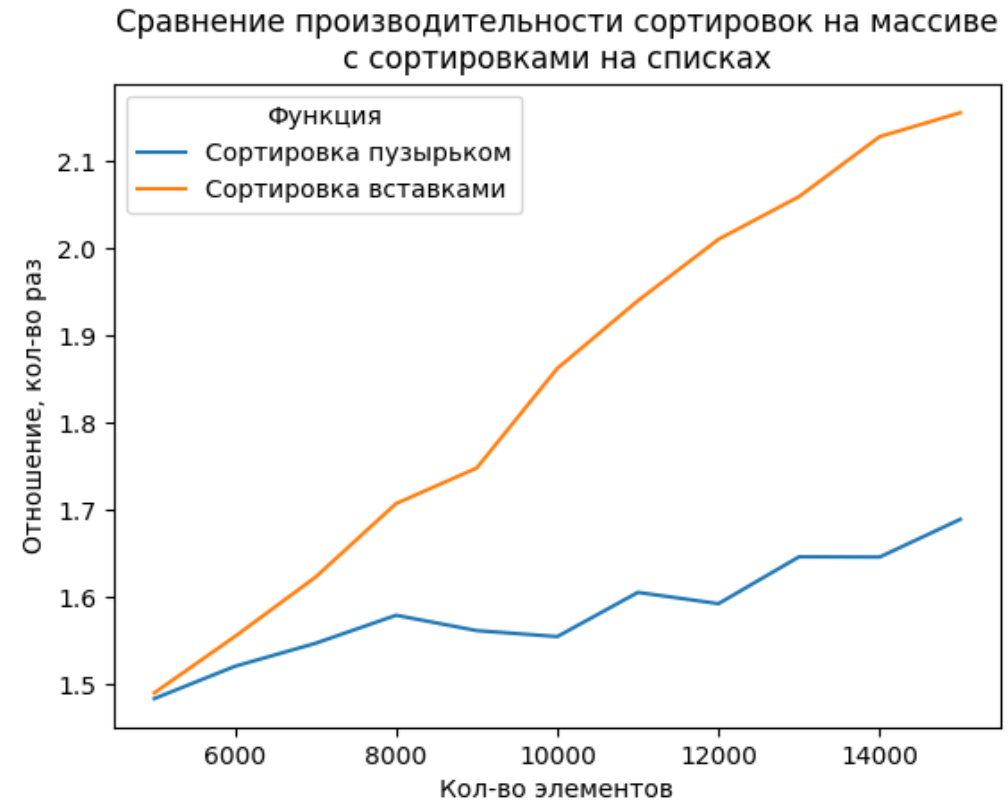
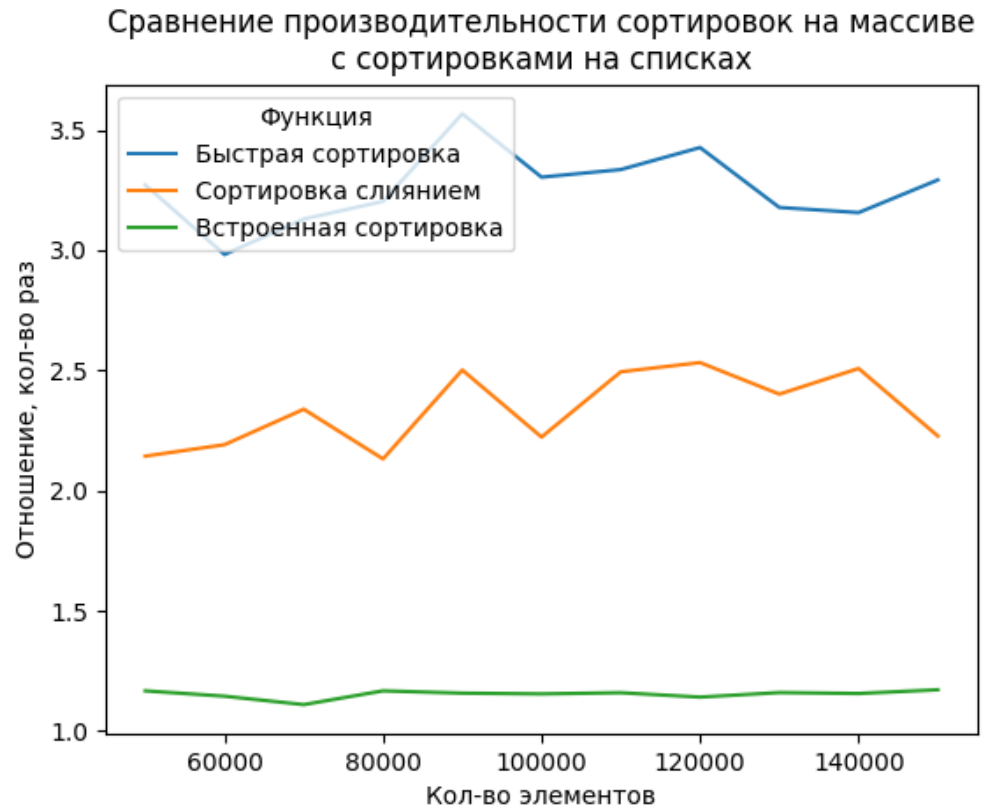
Сортировка пузырьком в 4 раза медленнее сортировки вставками.

Сравнение быстрых сортировок с «быстрой» сортировкой на массиве



Сортировки на массивах быстрее обеих сортировок на списках.
«Быстрая» сортировка медленнее сортировки слиянием.

Сравнение отношения производительности алгоритма на списке к производительности на массиве



Разница производительности на массиве и списке зависит от алгоритма.

Результаты

- «Быстрые» сортировки сильно быстрее «медленных».
- Самая быстрая среди быстрых на массивах и списках – сортировка слиянием, среди медленных – сортировка вставками.
- Сортировки на массивах быстрее сортировок на списках.