

Lớp: IT003.Q21.CTTN

BÁO CÁO KẾT QUẢ THỬ NGHIỆM

Sinh viên thực hiện: Đỗ Bách 25520110

Nội dung báo cáo: Cài đặt, so sánh và báo cáo kết quả thử nghiệm của các thuật toán sắp xếp

I. Kết quả thử nghiệm

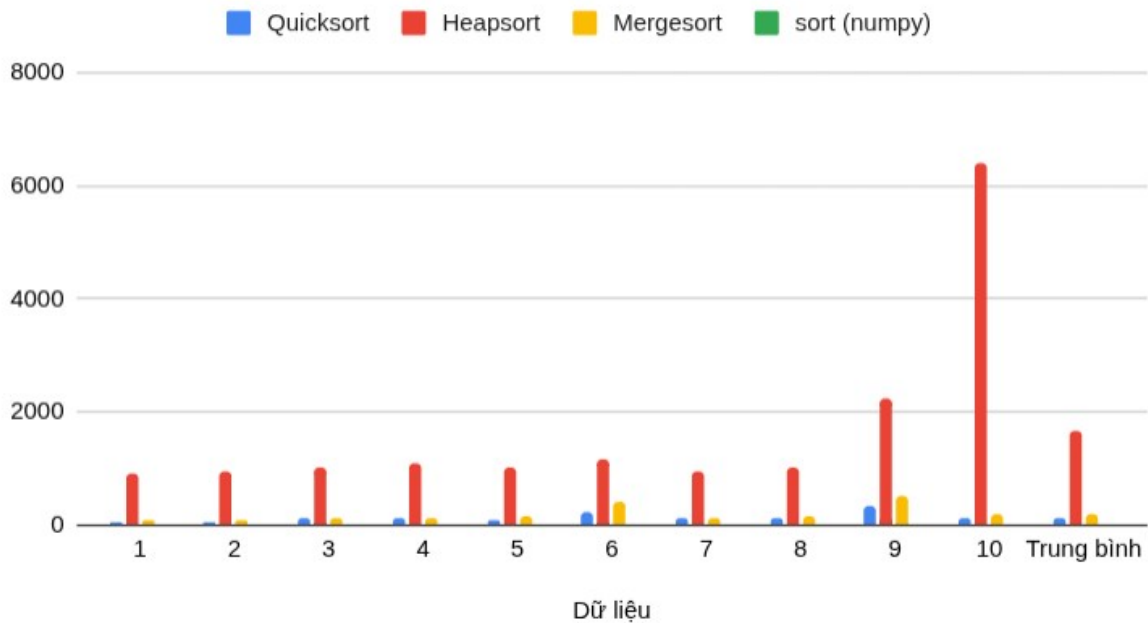
1. Bảng so sánh thời gian thực hiện.

Lưu ý: Tất cả đều được đưa về c sử dụng thư viện numba, njit.

Dữ liệu	Thời gian thực hiện (ms)			
	Quicksort	Heapsort	Mergesort	sort (numpy)
1	47	929	94	10
2	47	935	99	7
3	128	1011	140	9
4	108	1109	139	8
5	105	1003	144	7
6	240	1175	426	10
7	112	965	135	7
8	128	1035	168	10
9	334	2247	519	18
10	137	6393	190	9
Trung bình	138,6	1680,2	205,4	9,5

2. Biểu đồ (cột) thời gian thực hiện

Quicksort, Heapsort, Mergesort và sort (numpy)



II. Kết luận:

Đây là bài làm lại khi đã đưa tất cả thuật toán sắp xếp (cài đặt tay cũng như numpy sort) về dưới chung một định dạng là numpy. Theo em được biết và khuyến cáo thì numpy lấy code python và sau đó chuyển nó thành code c rồi chạy. Nhờ vậy mà nó có được một tốc độ tuyệt vời.

Tuy nhiên, do quy trình phức tạp ấy, các hành động thông thường (loop, while, if else từng phần tử) dựa trên mảng của numpy trở nên kém hiệu suất và thậm chí còn chậm hơn cả việc không sử dụng numpy.

Vì vậy, đoạn code của em đã phải sửa lại để cho đúng với tinh thần của numpy, đó chính là sử dụng thư viện numba với hàm wrapper njit. Khi chạy đoạn code python, nó sẽ lấy những hàm có wrapper này và compile code sang binary rồi chạy (giống như c). Khi ấy sức mạnh của numpy đã được khai thác tối đa

Nhìn vào kết quả thực nghiệm mới, ta có thể thấy rõ rằng thời gian đã được rút gọn hẳn so với việc cài đặt với list và ngôn ngữ python thông thường.

Heap sort là thuật toán chạy chậm nhất so với các thuật toán còn lại trong khi ấy quick sort và merge sort có thời gian chạy ngang ngửa nhau. Đáng chú ý, thư viện sort của numpy là chạy với thời gian ngắn nhất.

Điều này dẫn đến kết luận rằng tuy numpy có cơ chế giúp xử lý mảng nhanh hơn, ta phải sử dụng nó trong những trường hợp đặc biệt và kết hợp với những tài liệu trong document của nó. Và luôn luôn sử dụng thư viện có sẵn để giải quyết những bài toán lớn để tiết kiệm công sức và thời gian.

III. Trình bày thử nghiệm

Đường dẫn đến link github có file báo cáo và mã nguồn. <https://github.com/Do-Bach-2006/sort-algorithm-comparison>.

Do tập dữ liệu (10 triệu dòng) quá lớn, không thể tải lên github được nên em phải tải nó lên google drive

<https://drive.google.com/file/d/1VfrIGc1ZpCuwiiXyJ6wUQZbsD3SPWdPY/view?usp=sharing>