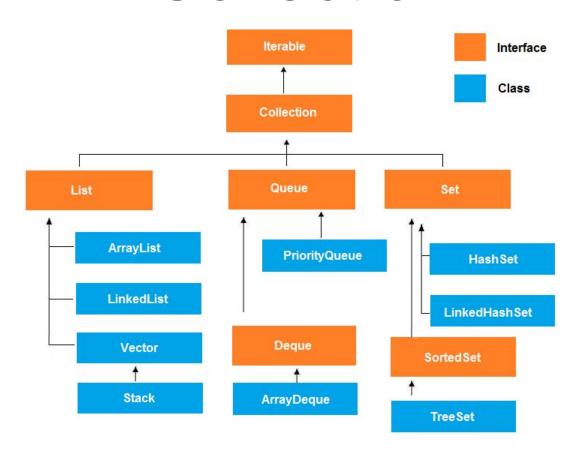
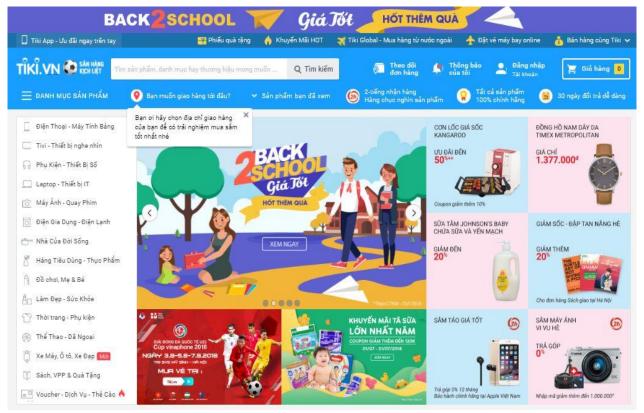
Collection



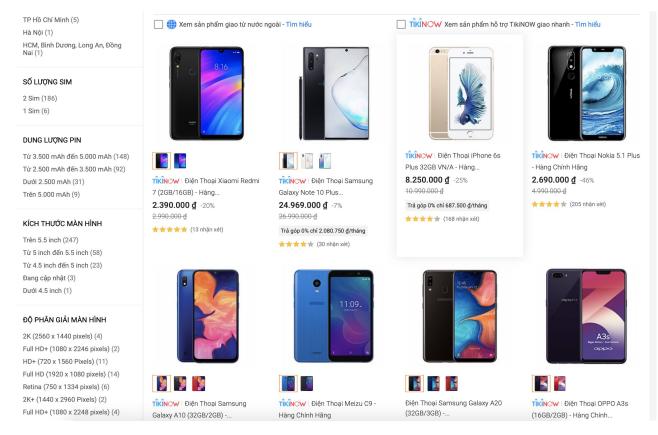
Collections

Collections trong java là một khuôn khổ cung cấp một kiến trúc để lưu trữ và thao tác tới nhóm các đối tượng. Tất cả các hoạt động mà bạn thực hiện trên một dữ liệu như tìm kiếm, phân loại, chèn, xóa,... có thể được thực hiện bởi Java Collections.



Set Interface

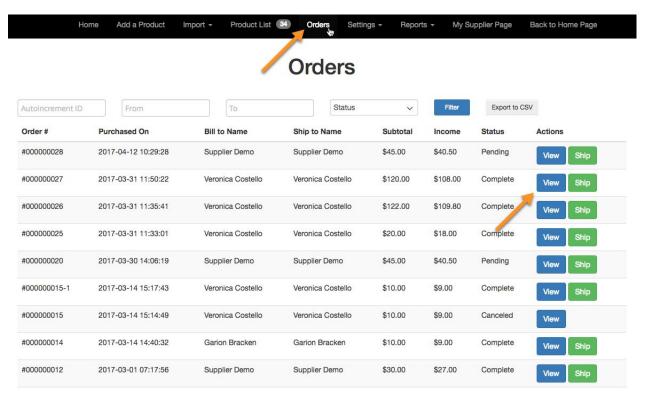
Set (tập hợp) là kiểu dữ liệu mà bên trong nó mỗi phần tử chỉ xuất hiện duy nhất một lần (tương tự như tập hợp trong toán học vậy) và tập hợp chưa được sắp xếp



Các điện thoại không được trùng nhau

List Interface

List (danh sách) là cấu trúc dữ liệu tuyến tính trong đó các phần tử được sắp xếp theo một thứ tự xác định. Cho phép các phần tử được trùng lặp nhau

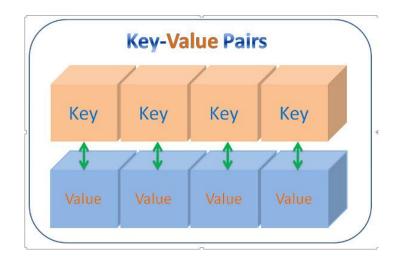


Ví du như Veronica

Map Interface

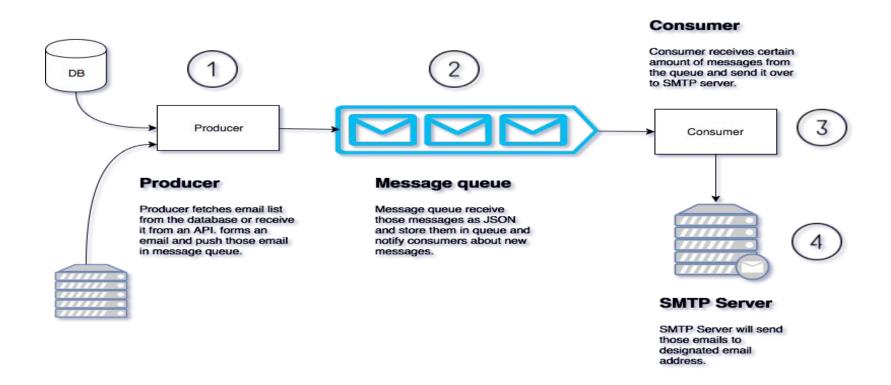
Map (đồ thị/ánh xạ) là kiểu dữ liệu cho phép ta quản lý dữ liệu theo dạng cặp key-value, trong đó key là duy nhất và tương ứng với 1 key là một giá trị value. Map Interface không kế thừa từ Collection Interface mà đây là 1 interface độc lập với các phương thức của riêng mình.





Queue Interface

Queue (hàng đợi) là kiểu dữ liệu nổi tiếng với kiểu vào ra FIFO (first-in-first-out hay vào trước ra trước)



Performance Big O

Khái niệm Big O hoạc với tên gọi khác trong tiếng Việt là "độ phức tạp của thuật toán" là thuật ngữ thường dùng để chỉ khoảng thời gian tiêu hao để chạy một thuật toán. Các lập trình viên thường sử dụng Big O như một phương tiện để so sánh mức độ hiệu quả của nhiều cách xử lý khác nhau cho cùng một vấn đề.

Khái niệm Big O có thể đúc kết thành một câu ngắn gọn như sau: Thời gian chạy nhanh như thế nào, còn tùy thuộc vào giá trị đầu vào -input, vì giá trị input sẽ lớn dần lên khi chương trình chạy.

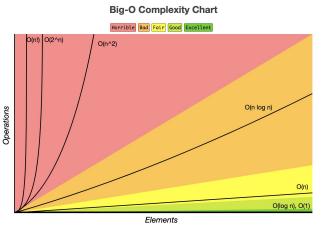
Performance Big O

Giá trị bất biến: Rõ ràng là Big O chỉ giải quyết vấn đề dựa trên tối ưu hóa phương thức code, bỏ qua các giá trị bất biến như đơn vị thời gian, size tiêu hao thực của bộ nhớ. Điều này rất bất lợi trong thời đại công nghệ phần cứng đang phát triển nhanh như vũ bão, tôi đang sử dụng một chiếc máy tính đời cổ để chạy một chương trình tốn đến 5 giờ, tôi miệt mài tốn công sức nghiên cứu tối ưu cách viết code để cải thiện chương trình chạy nhanh hơn 30 phút, thì cũng không bằng tôi ra tiệm sửa máy và nâng cấp phần cứng chiếc máy lên, khi đó máy tôi chạy chương trình nhanh hơn tận 4 tiếng.

Phương án tối ưu hóa code nóng vội: Đôi khi sử dụng phương thức Big O để tối ưu hóa code có thể tiết kiệm được thời gian chạy chương trình và giảm tiêu hao bộ nhớ, nhưng đồng thời sẽ khiến code khó đọc hoạc rất tốn thời gian suy nghĩ để viết. Đối với những lập trình viên trẻ, họ nên học cách viết code một cách tường minh, rõ ràng, dễ đọc, dễ hiểu và dễ bảo dưỡng dù chương trình họ viết ra có thể không được tiết kiệm tài nguyên như các lập trình viên lành nghề khác.

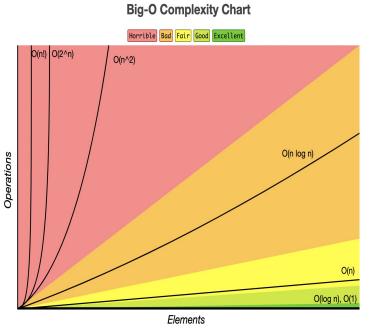
Nói như vậy không có nghĩa là các lập trình viên trẻ chả cần phải quan tâm gì đến Big O, vì bất kỳ một lập trình viên nào (dù trẻ, lành nghề, chuyên gia hay mới bước vào nghề) cũng luôn phải chú ý cân bằng các yếu tố thời gian chạy chương trình, thời gian chạy thuật toán, bộ nhớ tiêu hao, khả năng bảo dưỡng, tính tường minh của những dòng code.

Common Data Structure Operations



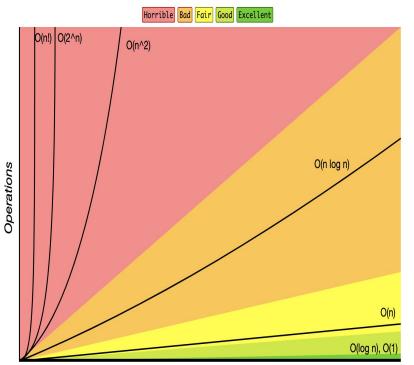
Data Structure		Space Complexity							
	Average				Worst		Worst		
	Access	Search	Insertion	Deletion	Access	Search	Insertion	Deletion	
<u>Array</u>	θ(1)	$\theta(n)$	$\theta(n)$	$\theta(n)$	0(1)	0(n)	0(n)	0(n)	0(n)
<u>Stack</u>	$\theta(n)$	θ(n)	θ(1)	Θ(1)	0(n)	0(n)	0(1)	0(1)	0(n)
<u>Queue</u>	Θ(n)	θ(n)	Θ(1)	θ(1)	0(n)	0(n)	0(1)	0(1)	0(n)
Singly-Linked List	Θ(n)	$\theta(n)$	Θ(1)	Θ(1)	0(n)	0(n)	0(1)	0(1)	0(n)
Doubly-Linked List	$\theta(n)$	θ(n)	Θ(1)	Θ(1)	0(n)	0(n)	0(1)	0(1)	0(n)
Skip List	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	0(n)	0(n)	0(n)	0(n)	0(n log(n))
Hash Table	N/A	Θ(1)	Θ(1)	θ(1)	N/A	0(n)	0(n)	0(n)	0(n)
Binary Search Tree	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	0(n)	0(n)	0(n)	0(n)	0(n)
Cartesian Tree	N/A	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	N/A	0(n)	0(n)	0(n)	0(n)
B-Tree	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	0(log(n))	0(log(n))	O(log(n))	O(log(n))	0(n)
Red-Black Tree	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	O(log(n))	0(log(n))	O(log(n))	O(log(n))	0(n)
Splay Tree	N/A	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	N/A	0(log(n))	0(log(n))	0(log(n))	0(n)
AVL Tree	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	0(log(n))	0(log(n))	0(log(n))	0(log(n))	0(n)
KD Tree	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	$\theta(\log(n))$	0(n)	0(n)	0(n)	0(n)	0(n)

Array Sorting Algorithms



Algorithm	Time Compl	exity	Space Complexity		
	Best	Average	Worst	Worst	
Quicksort	$\Omega(n \log(n))$	$\theta(n \log(n))$	0(n^2)	0(log(n))	
<u>Mergesort</u>	$\Omega(n \log(n))$	$\theta(n \log(n))$	0(n log(n))	0(n)	
<u>Timsort</u>	$\Omega(n)$	$\theta(n \log(n))$	0(n log(n))	0(n)	
<u>Heapsort</u>	$\Omega(n \log(n))$	$\theta(n \log(n))$	0(n log(n))	0(1)	
Bubble Sort	$\Omega(n)$	Θ(n^2)	0(n^2)	0(1)	
Insertion Sort	$\Omega(n)$	Θ(n^2)	0(n^2)	0(1)	
Selection Sort	Ω(n^2)	Θ(n^2)	0(n^2)	0(1)	
Tree Sort	$\Omega(n \log(n))$	$\theta(n \log(n))$	0(n^2)	0(n)	
Shell Sort	$\Omega(n \log(n))$	$\theta(n(\log(n))^2)$	0(n(log(n))^2)	0(1)	
Bucket Sort	$\Omega(n+k)$	$\theta(n+k)$	0(n^2)	0(n)	
Radix Sort	$\Omega(nk)$	θ(nk)	0(nk)	0(n+k)	
Counting Sort	$\Omega(n+k)$	$\theta(n+k)$	0(n+k)	0(k)	
Cubesort	$\Omega(n)$	$\theta(n \log(n))$	0(n log(n))	0(n)	

Big-O Complexity Chart



Elements

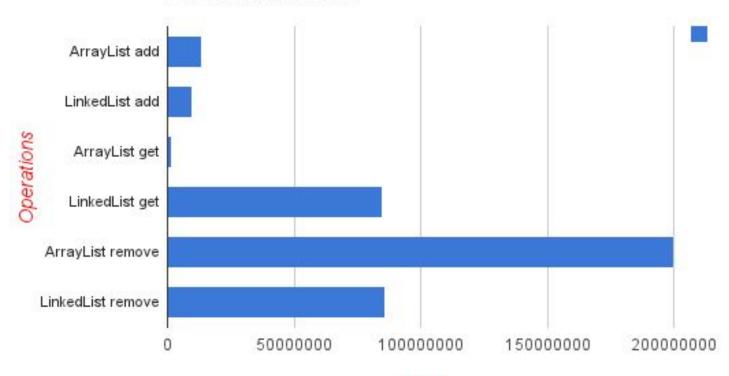
List	dd Remove Get Contains Ne	ext Data Structure
ArrayList	(1) O(n) O(1) O(n) O((1) Array
LinkedList	(1) 0(1) 0(n) 0(n) 0	(1) Linked List
CopyOnWriteArrayList	(n) 0(n) 0(1) 0(n) 0	(1) Array

Set	1	Add	 -	Remove		Contains		Next	1		Data Structure
HashSet	i	0(1)	i	0(1)	i	0(1)	i	0(h/n)	i	0(1)	Hash Table
LinkedHashSet	Ī	0(1)	ı	0(1)	ı	0(1)	1	0(1)	1	0(1)	Hash Table + Linked List
EnumSet	Ī	0(1)	ı	0(1)	ı	0(1)	1	0(1)	1	0(1)	Bit Vector
TreeSet	Ī	O(log n)	I	O(log n)	I	O(log n)	1	O(log n)	1	0(1)	Red-black tree
CopyOnWriteArraySet	Ī	0(n)	ı	0(n)	ı	0(n)	1	0(1)	1	0(1)	Array
ConcurrentSkipListSet	ī	O(log n)	i	O(log n)	ī	O(log n)	1	0(1)	1	0(n)	Skip List

Queue	1	Offer		I	Peak	1	Poll			Remove	1	Size Data Structure
	- -			۱-		-1			-		- -	
PriorityQueue	1	O(log n)	ı	0(1)	1	O(log r	1)		0(n)		O(1) Priority Heap
LinkedList	1	0(1)		l	0(1)	- [0(1)			0(1)	1	0(1) Array
ArrayDequeue	1	0(1)		ı	0(1)	-	0(1)			0(n)	1	0(1) Linked List
ConcurrentLinkedQueue	1	0(1)		ı	0(1)	-	0(1)			0(n)		O(n) Linked List
ArrayBlockingQueue	1	0(1)		ı	0(1)	-	0(1)			0(n)	1	0(1) Array
PriorirityBlockingQueue	1	O(log n)	l	0(1)	-	O(log r	١)		0(n)		O(1) Priority Heap
SynchronousQueue	1	0(1)		l	0(1)	-	0(1)			0(n)	1	0(1) None!
DelayQueue	1	O(log n)	ı	0(1)	-	O(log r	١)		0(n)	1	O(1) Priority Heap
LinkedBlockingQueue	1	0(1)		ı	0(1)	-	0(1)			0(n)		0(1) Linked List

Мар	I	Get	C	ontainsKey	1	Next		Data Structure
	1-				1-		1-	
HashMap	I	0(1)		0(1)	1	0(h / n)		Hash Table
LinkedHashMap	I	0(1)		0(1)	1	0(1)		Hash Table + Linked List
IdentityHashMap	I	0(1)		0(1)	1	0(h / n)		Array
WeakHashMap	I	0(1)		0(1)	I	0(h / n)		Hash Table
EnumMap	I	0(1)		0(1)	1	0(1)		Array
TreeMap	Ī	O(log n)	I	O(log n)	I	O(log n)		Red-black tree
ConcurrentHashMap	Ī	0(1)	I	0(1)	I	0(h / n)		Hash Tables
${\tt ConcurrentSkipListMap}$	I	O(log n)	1	O(log n)	I	0(1)		Skip List

Performance Test



Time

List Interface

List (danh sách) là cấu trúc dữ liệu tuyến tính trong đó các phần tử được sắp xếp theo một thứ tự xác định. List Interface định nghĩa các phương thức để tương tác với list cũng như các phần tử bên trong list. Tương tự như Set Interface, List Interface cũng được kế thừa và có đầy đủ các phương thức của Collection Interface.

- ArrayList: là 1 class dạng list được implement dựa trên mảng có kích thước thay đổi được.
- LinkedList: là một class dạng list hoạt động trên cơ sở của cấu trúc dữ liệu danh sách liên kết đôi (double-linked list)
- Vector: là 1 class thực thi giao diện List Interface, có cách thực lưu trữ như mảng tuy nhiên có kích thước thay đổi được, khá là tương tự với ArrayList, tuy nhiên điểm khác biệt là Vector là synchronized, hay là đồng bộ, có thể hoạt động đa luồng mà không cần gọi synchronize một cách tường minh
- Stack: cũng là 1 class dạng list, Stack có cách hoạt động dựa trên cơ sở của cấu trúc dữ liệu ngăn xếp (stack) với kiểu vào ra LIFO (last-in-first-out hay vào sau ra trước) nổi tiếng.

Queue Interface

Queue (hàng đợi) là kiểu dữ liệu nổi tiếng với kiểu vào ra FIFO (first-in-first-out hay vào trước ra trước), tuy nhiên với Queue Interface thì queue không chỉ còn dừng lại ở mức đơn giản như vậy mà nó cũng cấp cho bạn các phương thức để xây dựng các queue phức tạp hơn nhiều như priority queue (queue có ưu tiên), deque (queue 2 chiều), ... Và cũng giống như 2 interface trước, Queue Interface cũng kế thừa và mang đầy đủ phương thức từ Collection Interface.

Một số class về Queue thường sử dụng:

- LinkedList: chính là LinkedList mình đã nói ở phần List
- PriorityQueue: là 1 dạng queue mà trong đó các phần tử trong queue sẽ được sắp xếp.
- ArrayDeque: là 1 dạng deque (queue 2 chiều) được implement dựa trên mảng

TreeSet

Một số class thực thi Set Interface thường gặp:

- TreeSet
- HashSet: là 1 class implement Set Interface, mà các phần tử được lưu trữ dưới dạng bảng băm (hash table).
- EnumSet: là 1 class dạng set như 2 class ở trên, tuy nhiên khác với 2 class trên là các phần tử trong set là các enum chứ không phải object.

Map Interface

Map (đồ thị/ánh xạ) là kiểu dữ liệu cho phép ta quản lý dữ liệu theo dạng cặp key-value, trong đó key là duy nhất và tương ứng với 1 key là một giá trị value. Map Interface cung cấp cho ta các phương thức để tương tác với kiểu dữ liệu như vậy. Không giống như các interface ở trên, Map Interface không kế thừa từ Collection Interface mà đây là 1 interface độc lập với các phương thức của riêng mình.

Dưới đây là một số class về Map cần chú ý:

- TreeMap: là class thực thi giao diện Map Interface với dạng cây đỏ đen (Red-Black tree) trong đó các key đã được sắp xếp. Class này cho phép thời gian thêm, sửa, xóa và tìm kiếm 1 phần tử trong Map là tương đương nhau và đều là O(log(n))
- HashMap: là class thực thi giao diện Map Interface với các key được lưu trữ dưới dạng bảng băm, cho phép tìm kiếm nhanh O(1).
- EnumMap: cũng là 1 Map class nữa, tuy nhiên các key trong Map lại là các enum chứ không phải object như các dạng Map class ở trên.
- WeakHashMap: tương tự như HashMap tuy nhiên có 1 điểm khác biệt đáng chú ý là các key trong Map chỉ là các Weak reference (hay Weak key), có nghĩa là khi phần tử sẽ bị xóa khi key được giải phóng hay không còn một biến nào tham chiếu đến key nữa.