Collections with Uniqueness: Sets

Richard Warburton

Java champion, Author and Programmer

@richardwarburto www.insightfullogic.com

Sets are collections of distinct elements. There are no duplicates

Outline

Set Features

The Why and How of Sets

Hashcode and Equals

Understanding the contract behind HashSet

Set Implementations

Performance tradeoffs and features

Set Features

Hashcode and Equals


graphs, and hereby grants the
graphs for which the Client pays.
Signatures
Photographer's Signature

```
object.equals(other)
```

➔

```
object.hashCode() == other.hashCode()
```

Hashcode / Equals Contract

One way implication

Equality

It can be reference based or value based. Reference based just needs to inherit equals from Object. Value based requires a custom equals method.

```
result = 31 * result +
obj.hashCode();


// Arrays

Arrays.hashCode()


// Primitives (Java 8+)

Long.hashCode(longValue)


// Old Primitives

(int) (l ^ (l >>> 32))
Float.floatToIntBits(f);
```

- Combine hashcode information from each field
- IDE Can auto-generate
- Objects.hash() (Java 7+)
- ALWAYS use the same fields as equals()

Set Implementations

HashSet

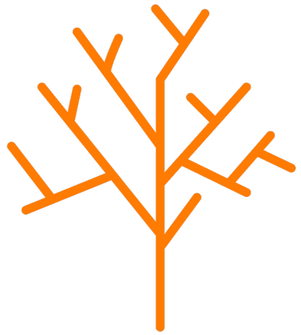
Based upon HashMap

Uses hashcode() and looks up location

Good General Purpose Implementation

Use by default

TreeSet



Based Upon TreeMap

Red/Black binary tree with defined sort order

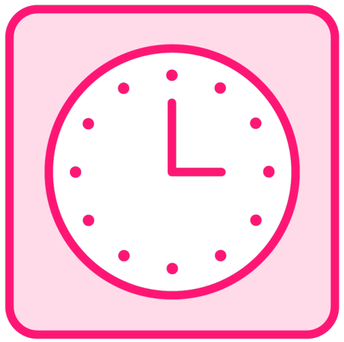
Provides Extra Features

Implements SortedSet and NavigableSet

Performance Comparison

	add	contains	next
HashSet	O(N) , Ω(1)	O(log(N)), Ω(1)	O(Capacity/N)
TreeSet	O(log(N))	O(log(N))	O(log(N))

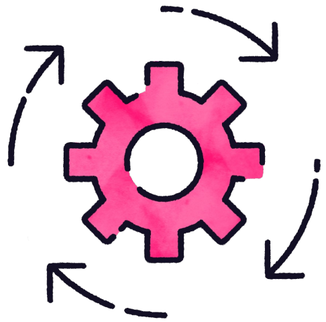
LinkedHashSet



When

Copying Set to modify

Deduping List or Queue



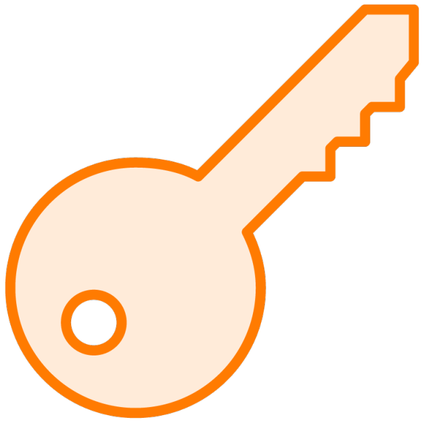
Maintains Order

Only Insertion

Overhead

Slower than HashSet, less memory than TreeSet

EnumSet



Keys are Enums

Faster and Low memory usage

Bitset Implementation

Only a single long if < 64 elements

SortedSet and NavigableSet

```
E first();

E last();


SortedSet tailSet(E fromElement);

SortedSet headSet(E toElement);

SortedSet subSet(E fromElement, E toElement);
```

SortedSet

Defines an order

No indexes, but subset views possible.

```
E lower(E e);
E higher(E e);

E floor(E e);
E ceiling(E e);

E pollFirst();
E pollLast();
```

NavigableSet

Extends SortedSet
Provides ways to move through the order
Implemented by TreeSet

Conclusion

Course Summary

Sets are a commonly used collection

Different implementations for different purposes

Remember to get the hashcode/equals contract correct


LINEX