Java `Map` and `Set` collections

- From last time: Comparator example
- Java `Set` container
  - idea
  - interface
- Java `Map` container
  - idea
  - interface
  - iterator
- concordance example

Announcements

- This week's lab based on an example we'll do in this Thurs lecture
- Sample MT 2 exams have been published.
- Reminders:
  - PA 3 due this Wed
  - MT 2 Tue. 4/4 – in THH 101 again

From last time: Additional `Comparator` example

- Problem: sort an array of **`Rectangle`**'s in increasing order by area.
- Do not implement your own sort method!

```
public static void sortIncrByArea(
                              Rectangle[] rects) {

    Arrays.sort(
```

Java Collections

- **`Collection`** is an interface in Java
- Linear collections:

  `ArrayList`, **`LinkedList`**, **`Stack`**, **`Queue`**

  – ordering of elements depended on order and type of insertion
- Two others today: **`Set`** and **`Map`**
  – ordering is determined internally by the class based on *value* of the element

Set ADT

(ADT = abstract data type)

Operations:

- add an element (no duplicate elements added)
- remove an element
- ask if an object is in the set
- list all the elements
  - (order of visiting depends on the kind of set created)

Set Examples

- Determine the number of unique words in a text file. (P16.1 from text)

- Spell-checker (Ex from Section 16.1 of text)

Java **`Set`** interface

- Two implementations:

```
Set<ElmtType> s = new HashSet<ElmtType>();
```

  - fastest. for when you don't care about order when iterating, or if you don't need to iterate.
  - **`ElmtType`** must support **`equals()`** and **`hashCode()`**

```
Set<ElmtType> s = new TreeSet<ElmtType>();
```

  - for when you need to visit element in sorted order.
  - **`ElmtType`** must implement **`Comparable`** (has **`compareTo`**)

- Normally use *interface* type for object reference. E.g.,

```
Set<String> uniqueWords =
                          new TreeSet<String>();
```

Java **`Set`** interface (cont.)

```
Set<String> uniqueWords =
        new TreeSet<String>();
```
creates empty set

```
uniqueWords.add("the");
```
if wasn't there, adds it and returns true,
o.w., returns false and set unchanged

```
uniqueWords.remove("blob");
```
if it was there, removes it and returns true,
o.w., returns false and set unchanged

```
uniqueWords.contains("the")
```
returns true iff "the" is in the set

```
size()   isEmpty()
```

Iterating over a Set

- **`Iterator`** is also an interface.
- Order elements visited depends on kind of Set involved.
- Can iterate over other Collections like we did with LinkedList. E.g.,

```
Set<String> uniqueWords = …;
...
Iterator<String> iter =
                    uniqueWords.iterator();
while (iter.hasNext()) {
    String word = iter.next();
    System.out.println(word);
          // or do something else with it
}
```

more about ElmtType

- best if it's an immutable type (e.g., **`String`**, **`Integer`**)
- Do not "mutate" element contents while in the Set:

```
Set<Point> setOfPoints = . . .
Iterator<Point> iter =
                    setOfPoints.iterator();
while (iter.hasNext()) {
    Point p = iter.next();
    p.translate(x, y); // BAD -- invalidates set
  }
```

Illustration of of mutating a value while it's in a Set

How many unique words in a file?

```
public static int numUnique(Scanner in) {
```

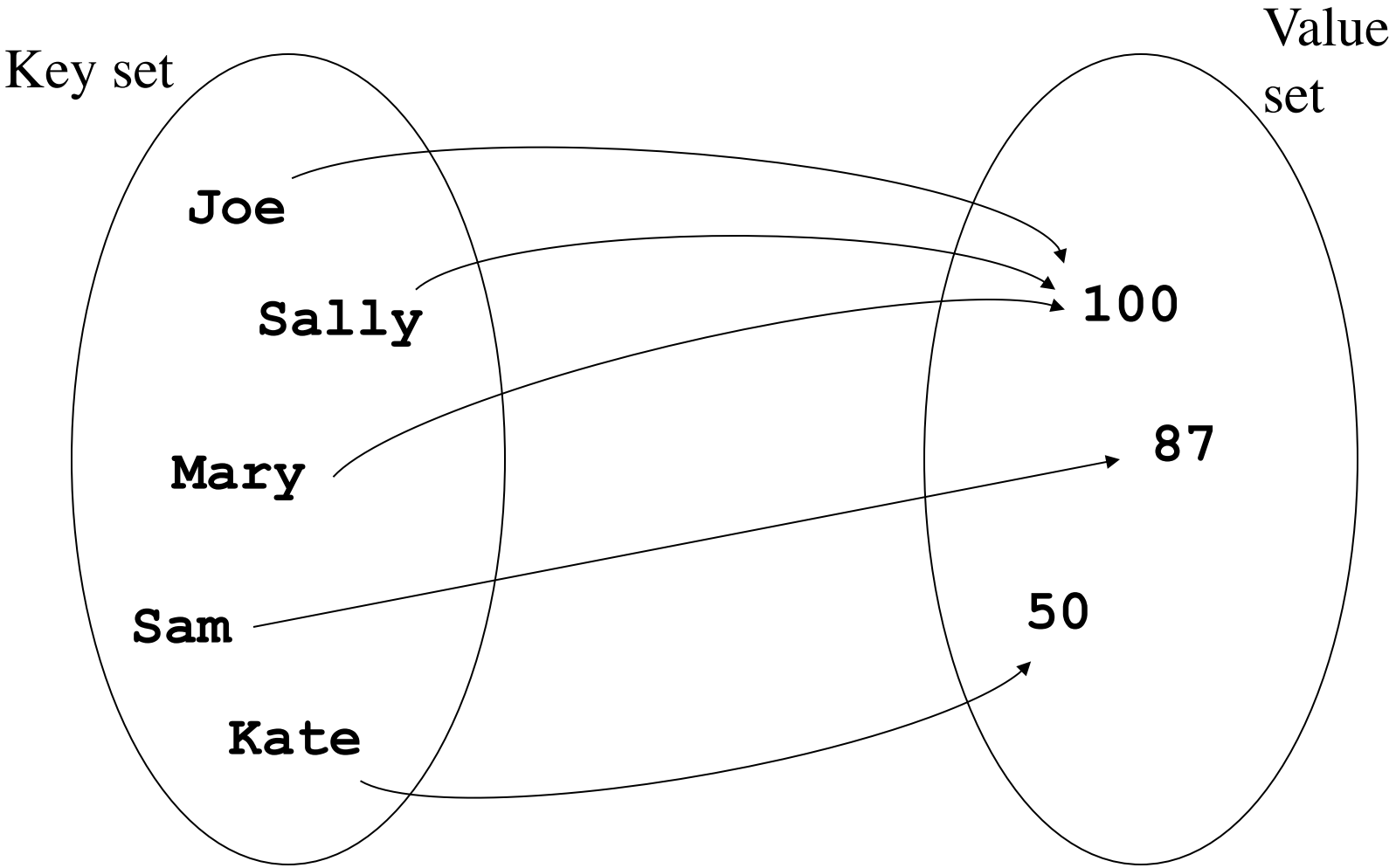
Map ADT

- A map stores a collection of (key,value) pairs
- keys are unique: a pair can be identified by its key

Operations:

- add a new (key, value) pair (called an *entry*)
- remove an entry, given its key
- lookup an entry, given its key
- list all the entries
  - (order of visiting depends on the kind of map created)

Example: map of students and their scores



Java **`Map`** interface

- Creation is same as Set, but *two* type parameters for generic class.

```
Map<KeyType, ValueType> map =
      new HashMap<KeyType,ValueType>();
```

  - fastest. for when you don't care about order when iterating, or if you don't need to iterate.
  - **`KeyType`** must support **`equals()`** and **`hashCode()`**

```
Map<KeyType, ValueType> map =
      new TreeMap<KeyType,ValueType>();
```

  - for when you need to visit element in sorted order by keys.
  - **`KeyType`** must implement **`Comparable`** (has **`compareTo`**)

Java **`Map`** interface (cont.)

- Create an empty map:

```
Map<String, Integer> scores =
            new TreeMap<String, Integer>();
```

- Note: **`put`** operation can be used in two ways:
- Suppose we do the two operations below in sequence:

```
scores.put("Joe", 98); // inserts
```

if key wasn't there, adds it and returns null,
o.w., returns the old value that went with this key

```
scores.put("Joe", 100); // updates
```

changes Joe's score to 100. if "Joe" hadn't been there before, this would have added him.

Java **Map** interface (cont.)

```
Map<String, Integer> scores =
        new TreeMap<String, Integer>();
```

```
scores.remove("Joe");
```

if key was there, removes it and returns the value that went with this key, o.w., returns null and map is unchanged

```
Integer jScore = scores.get("Joe");
```

return the value that goes with "Joe", or null if "Joe" is not in the map

Iterating over a Map

- A little different than Set or LinkedList.
- Suppose **`Map<String, Integer> scores`**
- Can iterate over all keys or all entries
- First get the "view" of the Map you need:
  - **`scores.keySet()`** returns the set of keys (type Set<String>)
  - **`scores.entrySet()`** returns a *set* whose elements are map entries (more details soon)
- Second, iterate over the set that was returned.

Iterating over all ***keys*** in a map

```
Map<String, Integer> scores =
        new TreeMap<String, Integer>();
. . .
Set<String> keySet = scores.keySet();
Iterator<String> iter = keySet.iterator();
while (iter.hasNext()) . . .
```

- Version without temp variable keySet:

```
Iterator<String> iter =
                  scores.keySet().iterator();
```

Iterating over all ***entries*** in a Map

- Using example map:

```
Map<String, Integer> scores;
```

- Reminder: **`scores.entrySet()`** returns a set of map entries.
- Elements of this set are type:

```
Map.Entry<String, Integer>
```

- Operations on a Map.Entry<K,V> entry:

```
entry.getKey()
entry.getValue()
entry.setValue(newVal)
```

Iterating over all entries in a Map (cont.)

- Example with **`Map<String, Integer> scores`**

```
Map<String, Integer> scores =
        new TreeMap<String, Integer>();
. . .
Iterator<Map.Entry<String, Integer>> iter =
            scores.entrySet().iterator();
while (iter.hasNext()) {
  Map.Entry<String, Integer> curr = iter.next();
  System.out.println(curr.getKey() + " ”
                              + curr.getValue());
}
```

for-each loop

- For some traversals we can use a for-each loop as a shortcut.
- General form (uses `for` keyword):

```
for (ElmtType elmt: collection) {
   do something with element
}
```

- Example with visiting all entries in a Map:

```
for (Map.Entry<String, Integer> entry:
                                        scores.entrySet()) {
    System.out.println(entry.getKey() + " "
                       + entry.getValue());
}
```

Final notes on Map interface

- Restrictions on KeyType in a Map

  (Same issue as with ElmtType of Set)
  - best if it's an immutable type (e.g., String, Integer)
  - Unsafe to "mutate" keys that are in a Map.
  - Entry's location in Map data structure depends on its key.
  - No restrictions on ValueType
- No iterator on Maps directly: have to use keySet() or entrySet() and iterate over resulting Set.

`Map` seen as an array

- Map ADT is sometimes called an *associative array*

```
System.out.println(scores.get("Joe"));
```

- ArrayList index syntax, but it's not random access
- But it's fast:
  - TreeMap: get, put, remove O(log n) each.
  - HashMap: get, put, remove O(1) each (!)
- E.g., Need an "array" indexed by a String?

. . . use a Map

Example: concordance

Problem: find the number of occurrences of each word in a text document.

- Why?
- (Variation also finds the page numbers or line numbers where those words occur in the document.)

Example: concordance (cont.)

- Similar to finding frequencies of student scores (from earlier in the semester):

```
// sample scores: 72 99 84 99 72 85 72 80
// scores are all in range [0..100]

int[] freq = new int[101];

for each score
  freq[score]++;
```

- Can we use an array in the same way for this problem?:

  Find the number of occurrences of each word in a text document.