

1 Portfolio Problem

Goals

This assignment consists of a small program that uses interfaces and classes either from Java's standard libraries, or from our earlier labs and assignments. The goal is to give you a bit of design freedom: You get to decide which parts of the standard libraries, or which classes and interfaces we already designed are the most suitable to use. If you design well, this assignment should be fairly straightforward.

The goal of the second part is to give you a practice in designing tests using the JUnit test tools and to create documentation in the style that allows you to produce *Javadoc* documentation for your program.

Hints

Some or all of the following interfaces and classes are likely to prove useful. In the *java.lang* package: *Comparable*, *Iterator*, *List*, *Map*, *Set*, *Collections*.

William Shakespeare

The Application

Have you ever wondered about the size of Shakespeare's vocabulary? For this assignment you will write a program that reads its input from a text file and lists the words that occur most frequently, together with a count of how many different words occur in the file. If this program were to run on a file that contains all of Shakespeare's works, it would tell you the approximate size of his vocabulary, and how often he uses the most common words.

Hamlet, for example, contains about 4542 distinct words, and the word "king" occurs 202 times.

The Problem

Start by downloading the file **WordCount.zip** and making an Eclipse project **WordCount** that contains these files. Run the project, to make sure you have all pieces in place. The main method is in the class *Examples*.

You are given the file **test.txt** that contains the entire text of *Hamlet* and a file **FileReader.java** that contains the code that generates the words from

the file `test.txt` one at a time, via an iterator.

Note: Here you will use the imperative Iterator interface that is a part of Java Standard Library. Make sure to look up the documentation for this interface and understand how it works.

The class *Examples* contain a skeleton of tests and the code that invokes the *main* method in the *FileReader* class that processes the input data.

Your tasks are the following:

1. Design the class *Word* to represent one word of Shakespeare's vocabulary, together with its frequency counter. The constructor takes only one *String* (for example the word "king") and starts the counter at one. We consider one *Word* instance to be equal to another, if they represent the same word, regardless of the value of the frequency counter. That means that you have to override the method *equals()* as well as the method *hashCode()*.
2. Design the class that implements the *Comparator* interface, so that the words can be sorted by frequencies. (Be careful!) When you are done, place this class definition as the last part of the class definition of the class *Word*. This is called an *inner class*.
3. Include in the class *Word* the method that allows you to increment the counter (using mutation), and a method *toString* that prints one line with the word and its frequency.
4. Design the class *WordCounter* that keeps track of all the words we have seen so far. It should include the following methods:

```
// records the Word objects generated by the given Iterator.  
void countWords (Iterator it) { ... }
```

```
// How many different Words has this WordCounter recorded?  
int words() { ... }
```

```
// Prints the n most common words and their frequencies.  
void printWords (int n) { ... }
```

Here are additional details:

5. *countWords* consumes an iterator that generates the words and builds the collection of the appropriate *Word* instances, with the correct frequencies.
6. *words* produces the total count of different words that have been consumed.
7. *printWords* consumes an integer *n* and prints the top *n* words with the highest frequencies (using the *toString* method defined in the class *Word*).

Test Design and Management

Of course, you need to test all methods as you are designing them. Design the tests in two stages:

1. First design the tests as we have done before, using the *tester.jar* and *interface.jar* test harness code.

This prepares us for a new way of running tests, namely using **JUnit** - Java's standard test framework.
2. **Introducing JUnit:** *To get the first taste of using JUnit, convert the tests for this problem to tests that use JUnit as follows:*

In the **File** menu select **New** then **JUnitTestCase**. When the wizard comes up, select to include the *main* method, the constructor, and the *setup* method. The tests for each of the methods will then become one test case similar to this one:

```
/**
 * Testing the method toString
 */
public void testToString(){
    assertEquals("Hello: 1\n", this.hello1.toString());
    assertEquals("Hello: 3\n", this.hello3.toString());
}
```

We see that *assertEquals* is basically the same as the *test* methods for our test harnesses, they just don't include the name of the test. Try to

see what happens when some of the tests fail, when a test throws an exception, and finally, make sure that at the end all tests succeed.

Note: JUnit uses Java *equals* method to compare two pieces of data for equality. Make sure your tests are designed to either compare the primitive results of methods, or, when comparing two instances of the class *Foo*, you have overridden the *equals* method in the class *Foo* to reflect your desired equality comparison.