

Programming Languages and Techniques (CIS120)

Lecture 32

April 8, 2016

Histogram Demo

Poll

Are you here today?

1. yes

Announcements

- HW8: Spellchecker
 - Available on the web site
 - Due: Tuesday
 - Parsing, working with I/O, more practice with collections
- Apply to be a TA!
 - CIS 120 TAs are THE BEST!
 - <http://www.cis.upenn.edu/~introtas>

Design Example: Histogram.java

A design exercise using `java.io` and
the generic collection libraries

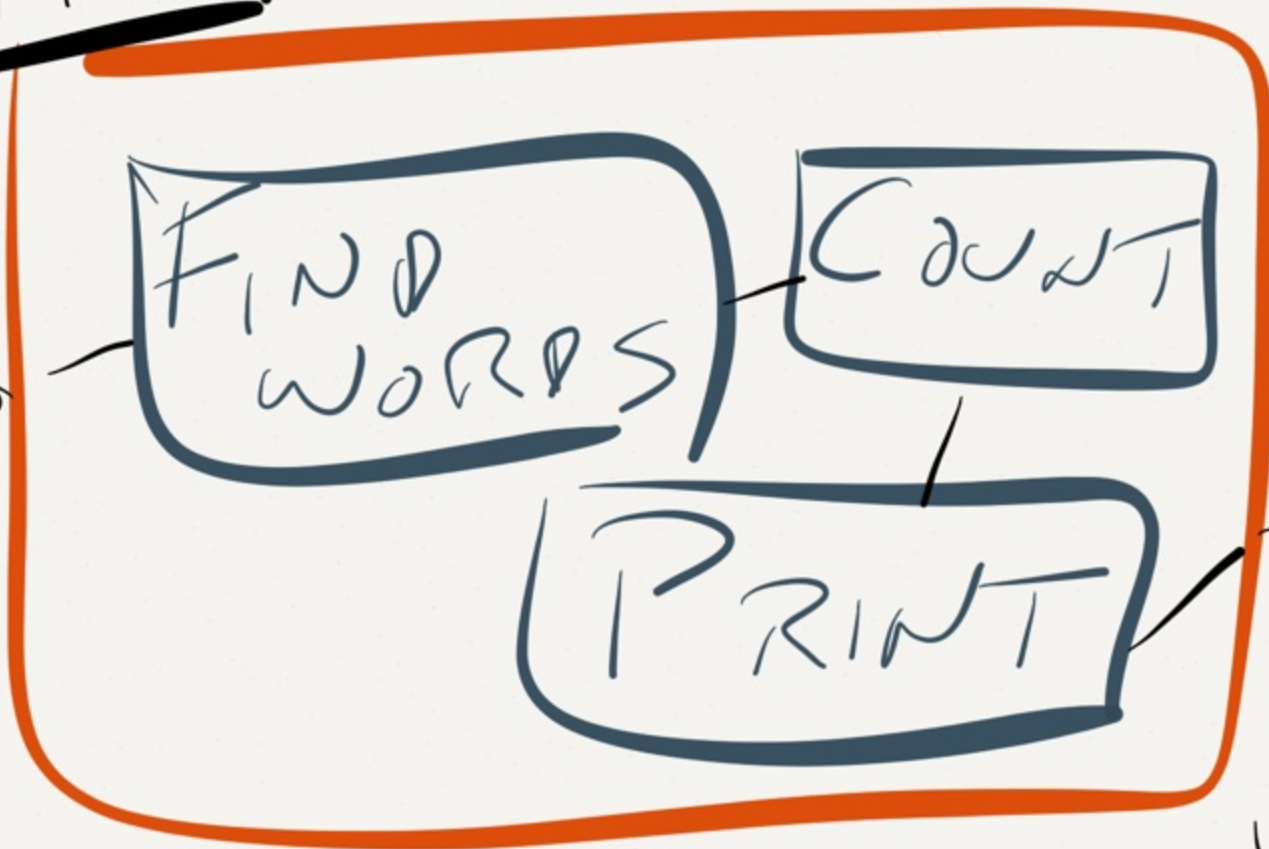
Problem Statement

Write a program that, given a filename for a text file as input, calculates the frequencies (i.e. number of occurrences) of each distinct word of the file. The program should then print the frequency distribution to the console as a sequence of “word: freq” pairs (one per line).

Histogram result:

The : 1	each : 1	line : 2	should : 1
Write : 1	file : 2	number : 1	text : 1
a : 4	filename : 1	occurrences : 1	that : 1
as : 2	for : 1	of : 4	the : 4
calculates : 1	freq : 1	one : 1	then : 1
command : 1	frequencies : 1	pairs : 1	to : 1
console : 1	frequency : 1	per : 1	word : 2
distinct : 1	given : 1	print : 1	
distribution : 1	i : 1	program : 2	
e : 1	input : 1	sequence : 1	

TEXT FILE



PRINTED HISTOGRAM

Reading Data

Which I/O class should we use to open the text file?

1. InputStream
2. FileInputStream
3. FileReader
4. BufferedReader

Decompose the problem

- Sub-problems:
 1. How do we iterate through the text file, identifying all of the words?
 2. Once we can produce a stream of words, how do we calculate their frequency?
 3. Once we have calculated the frequencies, how do we print out the result?
- What is the interface between these components?
- Can we test them individually?

How to produce a stream of words?

1. How do we iterate through the text file, identifying all of the words?

```
public interface Iterator<T> {  
    // returns true if the iteration has more elements  
    public boolean hasNext();  
    // returns the next element in the iteration  
    public T next();  
    // Optional: removes last element returned  
    public void remove();  
}
```

- **Key idea:** Define a class (WordScanner) that implements this interface by reading words from a text file.

Interactive Demo

Histogram.java

2. Once we can produce a stream of words, how do we calculate their frequency?
3. Once we have calculated the frequencies, how do we print out the result?

Interactive Demo

WordScanner.java