SUTD 50.001 Introduction to Information Systems and Programming

Problem set 1 – Part B (Week 1 to 3)

Your task is to implement the code to meet the **specification written as comments in the starting code.**

The problems are presented in a logical order, but you will receive credit for any parts of the problem set completed, even if you did not complete the prerequisites for that part. So, work ahead if you are stuck on any component of the problem set.

Do not change the signatures or specifications of any methods, classes, or packages that we have provided you. Your code will be tested automatically, and will break our testing suite if you do so.

This problem set is designed so that it can be done with completely stateless functions. All of the functions we have provided you to implement are labeled static. You shouldn't need to introduce class variables or non-static functions in this problem set.

Note:

- **Work on the problems in Android studio, and this includes writing code for the test cases. The starter code is provided separately.**
- **Vocareum is for submission only. Clicking the Submit button merely submits your code for our grading – no test cases will be provided there.**

Overview

The theme of this project is to find English words in the digits of Pi. To do this, we'll:

1. Compute the fractional digits of Pi
2. Convert the digits of Pi into a numeric base more suitable for word-finding -- i.e., base 26.
3. Transform the digits of Pi into an alphabetic String of letters
4. Find words in that particular encoding of Pi

After the basic version of the code works, we'll work on improving it to get better word coverage -- changing the mapping of digits to letters so that you're more likely to find interesting words.

Your workflow for this problem set should be:

1. Carefully look at the specification of the method you are looking to implement. The specifications are written as comments in the starting code.
2. Write the actual method we specified.

Pi Generation

This part of the problem set will generate an arbitrary number of digits of Pi, so we have data to search through. We will be implementing the Bailey–Borwein–Plouffe formula. Because this algorithm lets us generate arbitrary digits of Pi, it is much easier to test than other Pi generation methods.

Note that the BBP algorithm returns digits of Pi in base-16. Throughout Computer Science, base-16 is commonly referred to as hexadecimal (or hex). Hex is usually expressed using 0-9 and A-F to represent digits, with A = 10, B = 11, ..., F = 15. You'll see that terminology used in the problem set.

We have provided you with almost all of the implementation of the algorithm, the bulk of which can be found in PiGenerator.java.

For this part of the problem, you'll have to test and implement computePiInHex() and powerMod() in PiGenerator.java. Assume that the implementation of piTerm() and piDigit() we've provided you are correct.

a. [10 points] powerMod()
Implement powerMod() in PiGenerator.java.

b. [10 points] computePiInHex()
Implement computePiInHex() in PiGenerator.java. Note that this function should only return the fractional digits of Pi, and not the leading 3.

Executing Main.java now should print you some of the hexadecimal digits of Pi. You should verify that the output is what you expect. You can modify PI_PRECISION at the top of the file to change how many digits of Pi to generate; generating less digits will be faster.

Transforming Pi

In order to find words in the digits of Pi, we'll first need to convert the hex digits of Pi into something more useful. As a first try, we'll convert the hex digits into base-26, then do the straightforward mapping of numbers to letters. This part of the problem set will implement BaseTranslator.convertBase()

c. [20 points] convertBase()

Implement convertBase() in BaseTranslator.java, verifying that the tests you've written for it pass.

Executing Main.java should print you the base-26 digits of Pi. Verify that the numbers are what you expect.

Converting Pi to Characters

Now that we have Pi in base-26, it is a straightforward task to convert it to a string of letters. This part of the problem set will implement DigitsToStringConverter.convertDigitsToStrings().

d. [20 points] convertDigitsToString()
Implement convertDigitsToString() in DigitsToStringConverter.java

Executing Main.java should print you the translation of base-26 Pi into a-z characters.

Finding Words
Now that we have an alphanumeric string, we want to find words in it.

e. [20 points] getSubstrings()
Implement getSubstrings() in WordFinder.java, verifying that it conforms to the listed specification, and that all of your tests pass.

Executing Main.java now should show you a list of the words that were found in the digits of Pi! It should also tell you what percentage of words were found from the word list included in the assignment.

Alphabet Generation Revisited

As you can see, we don't do very well with finding most words in the digits of Pi. Part of the reason is that the alphabet we're using is not very smart; "z"s occur with roughly the same frequency as "e"s. This problem will explore one way to improve our implementation.

We'll use the word list that is included in the code to take a guess at how often each character occurs relative to the other characters, and we'll weigh the output alphabet we're translating with in favor of more frequently occurring characters.

f. [20 points] generateFrequencyAlphabet()
Implement generateFrequencyAlphabet() in AlphabetGenerator.java

Executing Main.java now should show a list of words that were found in the digits of Pi using the alternative alphabet. The coverage you should get for this implementation should be higher than in the basic one.

Before You're Done...

Double check that you didn't change the signatures of any of the code we gave you.

Make sure the code you implemented doesn't print anything to System.out. It's a helpful debugging feature, but writing output is a definite side effect of methods, and none of the methods we gave you to write should produce any side effects.

Make sure you don't have any outdated comments in the code you turn in. In particular, get rid of blocks of code that you may have commented out when doing the pset, and get rid of any TODO comments that are no longer TODOs.

Make sure your code compiles, and all the methods you've implemented pass all the tests that you added.

Does your code compile without warnings? In particular, you should have no unused variables, and no unneeded imports.

Some notes

We commonly use the decimal number system (base-10). Hence

$$45.123 = 4 \times 10^1 + 5 \times 10^0 + 1 \times 10^{-1} + 2 \times 10^{-2} + 3 \times 10^{-3}$$

There are other bases that are used in computing, you may have heard of:

- binary numbers (base-2):
  $1011_2 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$
- hexadecimal numbers (base-16):
  $4F.2A_{16} = 4 \times 16^1 + 15 \times 16^0 + 2 \times 16^{-1} + 10 \times 16^{-2}$

You can have numbers in any base you like. To convert 2.1_{10} to base 3 means solving the equation:

$$2 \times 10^0 + 1 \times 10^{-1} = x_0 \times 3^0 + x_{-1} \times 3^{-1} + x_{-2} \times 3^{-2} + \ldots$$

The number of terms on the RHS depends on the precision that you desire.