

DNA Sequences (Due 22 Mar 2013)

DNA or deoxyribonucleic acid is a nucleic acid that contains genetic information. It is responsible for propagation of inherited traits. DNA is organized as two complementary strands that Watson and Crick called the *Double Helix*. Each strand is built out of nucleotides called bases of which there are four - adenine (A), thymine (T), cytosine (C), and guanine (G). The bases of the two complementary strands that make up the DNA pair up in this order: A+T, T+A, C+G, and G+C. Strands have directionality and the sequence order does matter. Genetic information is determined by the sequence of bases along the strand.

DNA has played an important role in research in computer science. For example research in string searching algorithms has been motivated by finding sequences in DNAs. For the present assignment, we are interested in finding the longest common base sequence in two DNA strands. Each strand is represented by the sequence of letters A, T, C, and G. For the two strands *ACTG* and *TGCA* the longest common sequence is *TG*. It is quite possible for two strands not to have any common sequence (a sequence of 1 base does not count). Also there could be two or more common sequences that have the same longest length.

In your program you will open a text file called [dna.txt](#). The first line of data is an integer number n that gives the number of pairs of DNA to follow. You will read one pair of DNA strings at a time. The maximum length of each string is 80 characters. Assume that each string consists only of characters 'A', 'T', 'C' and 'G'. It is acceptable if a string is in lower case or is in mixed upper and lower case. Convert both strings to upper case. Print out the longest common sequence(s) for the two strings. If there is more than one longest common sequence then print each of those sequences on separate lines. The sequences should be left aligned. If there is no common sequence your program should output *No Common Sequence Found*.

Sample output session would look like:

```
Longest Common Sequences

Pair 1: TCG

Pair 2: TGAT
      GGAC

Pair 3: No Common Sequence Found
```

We will be looking at good documentation, descriptive variable names, and adherence to the coding convention mentioned below. Your file `DNA.java` will have the following header:

```
/*
File: DNA.java

Description:

Student Name:

Student UT EID:

Course Name: CS 312

Unique Number:
```

```
Date Created:
```

```
Date Last Modified:
```

```
*/
```

You will follow the standard Java [Coding Conventions](#). You can either view the HTML page or download the PDF or Postscript and print it out. There is a modification that I would like to make to the standard coding conventions. Please align the opening and closing braces vertically so that you can easily make out the blocks of code. For example:

```
Do this:
if ( x > 5 )
{
    a = b + c;
}
```

```
Not this:
if ( x > 5 ) {
    a = b + c;
}
```

Use the [turnin](#) program to submit your **.java** file. The proctors should receive your work by 11 PM on Friday, 22 Mar 2013. There will be substantial penalties if you do not adhere to the guidelines.

- You must submit the .java file and not the .class file.
- Your .java file should have the header with the proper documentation.
- You should be submitting your .java file through the web based *turnin* program. We will not accept files e-mailed to us.
- Compile and run your code on the command line.
- Your code must compile before submission.
- Here is the [Grading Criteria](#).