nscc

Assignment 1: Programming for AI

Question 1: Phone Words

Phone Words

Create a Python application that will convert a *PhoneWord* or Vanity Number (https://en.wikipedia.org/wiki/Phoneword) to a set of numbers. An example would be taking 1-800-HOLIDAY and getting 18004654329.

The image below shows the keypad you will be emulating:

Requirements

Given any PhoneWord:

- Ignore any non -alphanumeric characters: (), -, etc.
- Keep any existing digits
- Resolve a letter to a number according to the keypad image

Question 2: Bioinformatics 101

Step 1 - Convert DNA Sequence to corresponding Codon Sequence

DNA code is comprised of four letters: G, T, A, and C. In a strand of DNA, each triplet of these letters is a called a *codon*. Each codon represents an amino acid. Your first task in this tech check is to convert a DNA sequence – stored as a string – into the corresponding sequence of codons.

So, for example, the following DNA sequence string: GCTCGTAATGATTGT should be converted into the following codon sequence: **["GCT","CGT","AAT","GAT","TGT"]**.

Step 2 - Convert your codon sequence into an amino acid sequence

Once you've translated the DNA sequence into a codon sequence, your next job is to translate that codon sequence into an amino acid sequence. Imagine doing that by hand using the table below.

There are 21 amino acids. The DNA codons that represent them and their abbreviations are specified in the table below.

Your job is to programatically perform the conversion using a data file that you create that represents the mappings between codon triplets and amino acids. You can use any file format you prefer.

Abbr.	DNA Codons	Amino Acid
Ala	GCT, GCC, GCA, GCG	Alanine
Arg	CGT, CGC, CGA, CGG, AGA, AGG	Arginine
Asn	AAT, AAC	Asparagine
Asp	GAT, GAC	Aspartic Acid
Cys	TGT, TGC	Cysteine
Gln	CAA, CAG	Glutamine
Glu	GAA, GAG	Glutamic acid
Gly	GGT, GGC, GGA, GGG	Glycine
His	CAT, CAC	Histidine
Ile	ATT, ATC, ATA	Isoleucine
Leu	CTT, CTC, CTA, CTG, TTA, TTG	Leucine
Lys	AAA, AAG	Lysine
Met	ATG	Methionine
Phe	TTT, TTC	Phenylalanine
Pro	CCT, CCC, CCA, CCG	Proline
Pyl	UAG	Pyrrolysine
Ser	TCT, TCC, TCA, TCG, AGT, AGC	Serine
Sec	UGA	Selenocysteine
Thr	ACT, ACC, ACA, ACG	Threonine
Trp	TGG	Tryptophan
Tyr	TAT, TAC	Tyrosine

Abbr.	DNA Codons	Amino Acid
Val	GTT, GTC, GTA, GTG	Valine

Submission Instructions

When your programs are complete, move the folders containing each program into one master folder and zip up that master folder using standard, Windows built-in, pkzip (right-click folder and Send To->compressed file). Submit the zip file on Brightspace.