Exercici 6: Algorismes per text.

Lliurament:

UN ÚNIC FITXER (exercici6.py) QUE CONTINGUI EL CONJUNT DE FUNCIONS QUE S'HAN IMPLEMENTAT.

Criteris d'avaluació (per ordre d'importància):
1.- El programa ha de donar un resultat correcte: 50% de la nota.
2.- Ús adequat del llenguatge (fer servir if/while correctament, fer servir la col·lecció adequada, etc.): 30% de la nota.
3.- Bon estil de programació (fer una interfase d'usuari adequada, comentaris, etc.): 20% de la nota.

- Una **seqüència genètica** és un *string* format per caràcters d'un alfabet de quatre lletres: A, T, G i C, que corresponen a les macromolècules base de l'ADN. Un gen és una seqüència genètica que conté la informació necessària per construir una proteïna. La unió de tots els gens constitueixen el genoma.
  Cada cèl·lula produïda pel cos rep una còpia del genoma, però a vegades, aquesta còpia és lleugerament "defectuosa". Els "defectes" van des de la substitució d'una base per una altra fins a la pèrdua d'un substring de la seqüència.
  Baixeu-vos el fitxer HUMAN-DNA.txt al vostre directori Python. Aquest fitxer conté una part del ADN del cromosoma 2 humà.
  Feu una funció, anomenada **dna**, basada en l'algorisme de **Levenshtein**, que busqui dins de cada una de les linies del fitxer anterior les següents seqüències genètiques i digui on les ha trobat i amb quina distància:

  AGATACATTAGACAATAGAGATGTGGTC
  GTCAGTCTGGCCTTGCCATTGGTGCCACCA
  TACCGAGAAGCTGGATTACAGCATGTACCATCAT

  En les aplicacions bioinformàtiques, els costos de les operacions d'edició són lleugerament diferents de les que hem vist fins ara:
  - Per un salt (al patró o al text): 2
  - Per una substitució: 1
  - Quan hi ha corrrespondència: 0

  Adapteu la funció a aquests costos.
  La funció no ha de tenir cap tipus d'entrada de part de l'usuari, i la sortida ha de ser la següent:

```
El patró AGATACATTAGACAATAGAGATGTGGTC es troba a la línia X1, posició Y1, del
cromosoma 2 humà, i la seva distància d'edició és Z1.
El substring del cromosoma humà més semblant és S1.
El temps de càlcul ha estat T1.

El patró GTCAGTCTGGCCTTGCCATTGGTGCCACCA es troba a la línia X2, posició Y2,  del
cromosoma 2 humà, i la seva distància d'edició és Z2.
El substring del cromosoma humà més semblant és S2.
El temps de càlcul ha estat T2.

El patró TACCGAGAAGCTGGATTACAGCATGTACCATCAT es troba a la línia X3, posició Y3, del
cromosoma 2 humà, i la seva distància d'edició és Z3.
El substring del cromosoma humà més semblant és S3.
El temps de càlcul ha estat T2.
```

Recordatori de teoria: *El càlcul de de la distancia d'un patró al substring més semblant d'un text es pot fer amb* ***l'algorisme de Levenshtein****. L'única diferència és que s'ha d'inicialitzar la* ***primera fila amb zeros (=considerar que podem inserir tants espais en blanc al davant del patró com sigui necessari)*** *i la distancia d'edició serà el valor mínim de l'última fila de la matriu de costos. Per exemple, si el patró és "aba" i el tex tés "c abba c", la matriu és:*

		c	.	a	b	b	a	.	c
	0	0	0	0	0	0	0	0	0
a	1	1	1	**0**	1	1	0	1	1
b	2	2	2	1	**0**	1	1	1	2
a	3	3	3	2	1	**1**	**1**	2	2