IDC306: Biocomputing
Assignment-11
Date: 19 Mar 2024

Q1. Parse the **uniprotSnip.fasta** to extract information and store in dictionary. The information to store is:
UniqueIdentifier, EntryName, OrganismName, GeneName, ProteinExistence, ProteinSequence

----
>db|UniqueIdentifier|EntryName ProteinName OS=OrganismName OX=OrganismIdentifier GN=GeneName PE=ProteinExistence SV=SequenceVersion
ProteinSequence

Details

All entries will have:
db: sp or tr
UniqueIdentifier: unique id of a protein sequence
ProteinName: could be of multiple words
OX: Essentially taxonomic identifier
PE
SV
Entries may not have GN entry

Q2. Parse the **unitprotSnip.txt** to extract. Write a function, which will extract 'ID, AC, DR (multiple cross-reference database), SQ, (length and MW)'