Homework 10

Rohit Kamat rgk359

This homework is due on April 18, 2017 at 7:00pm. Please submit as a PDF file on Canvas. Before submission, please re-run all cells by clicking "Kernel" and selecting "Restart & Run All."

Problem 1 (2 points): Using Biopython and the Pubmed database, calculate the total number of papers that Dr. Wilke has published from 2011-2016 (inclusive, so that's 6 years total).

HINTS: Dr. Wilke will always appear as "Wilke CO" in the Pubmed database. Also, make sure to set the `retmax` argument to at least `60` in `Entrez.esearch()` so that you retrieve all of the papers.

```
In [1]:  # You will need Entrez and Medline to solve this problem
         from Bio import Entrez, Medline

         Entrez.email = "rohkamat@yahoo.com"

         # Your code goes here
         handle = Entrez.esearch(db="pubmed", term="Wilke CO[Author] AND 2011:2016[Da
         record = Entrez.read(handle)
         handle.close()
         pmid_list = record["IdList"]
         print("Wilke has been in", len(pmid_list), "published papers.")
```

```
Wilke has been in 56 published papers.
```

Problem 2 (4 points): From the years 2011-2016 (inclusive), in which journals did Dr. Wilke's papers appear? Print out all journals, in alphabetical order, making sure you don't print the same journal name twice.

Hint: You can use the optional `key` argument in the sort function to make sure the sort is not case-sensitive.

```
In [5]: from Bio import Medline
        handle = Entrez.efetch(db="pubmed", id=pmid_list, rettype="medline", retmode
        records = Medline.parse(handle)
        title_list= []
        for record in records:
            title_list=list(title_list)
            title_list.append(record['JT'].title())

        sorted(title_list)
        title_list= set(title_list)


        handle.close()

        title_list
```

```
Out[5]: {'Aids Research And Human Retroviruses',
         'Biology Direct',
         'Bmc Evolutionary Biology',
         'Elife',
         'Epidemics',
         'Evolution; International Journal Of Organic Evolution',
         'Genetics',
         'Genome Biology And Evolution',
         'Ieee/Acm Transactions On Computational Biology And Bioinformatics',
         'Infection, Genetics And Evolution : Journal Of Molecular Epidemiology A
        nd Evolutionary Genetics In Infectious Diseases',
         'Interdisciplinary Perspectives On Infectious Diseases',
         'International Forum Of Allergy & Rhinology',
         'Journal Of Molecular Evolution',
         'Journal Of Molecular Microbiology And Biotechnology',
         'Journal Of The Royal Society, Interface',
         'Journal Of Virology',
         'Molecular Biology And Evolution',
         'Nature Reviews. Genetics',
         'Peerj',
         'Philosophical Transactions Of The Royal Society Of London. Series B, Bi
        ological Sciences',
         'Physical Biology',
         'Plos Biology',
         'Plos Computational Biology',
         'Plos Genetics',
         'Plos One',
         'Plos Pathogens',
         'Proceedings Of The National Academy Of Sciences Of The United States Of
        America',
         'Protein Science : A Publication Of The Protein Society',
         'Proteins',
         'Science (New York, N.Y.)',
         'The Journal Of General Virology',
         'The Laryngoscope',
         'Virus Evolution'}
```