

Welcome to BST 281 Lab 6

1 Mar, 2018

Mike MacArthur

macarthur@g.harvard.edu

Office Hours: Fridays 2-3p

Kresge Student Lounge

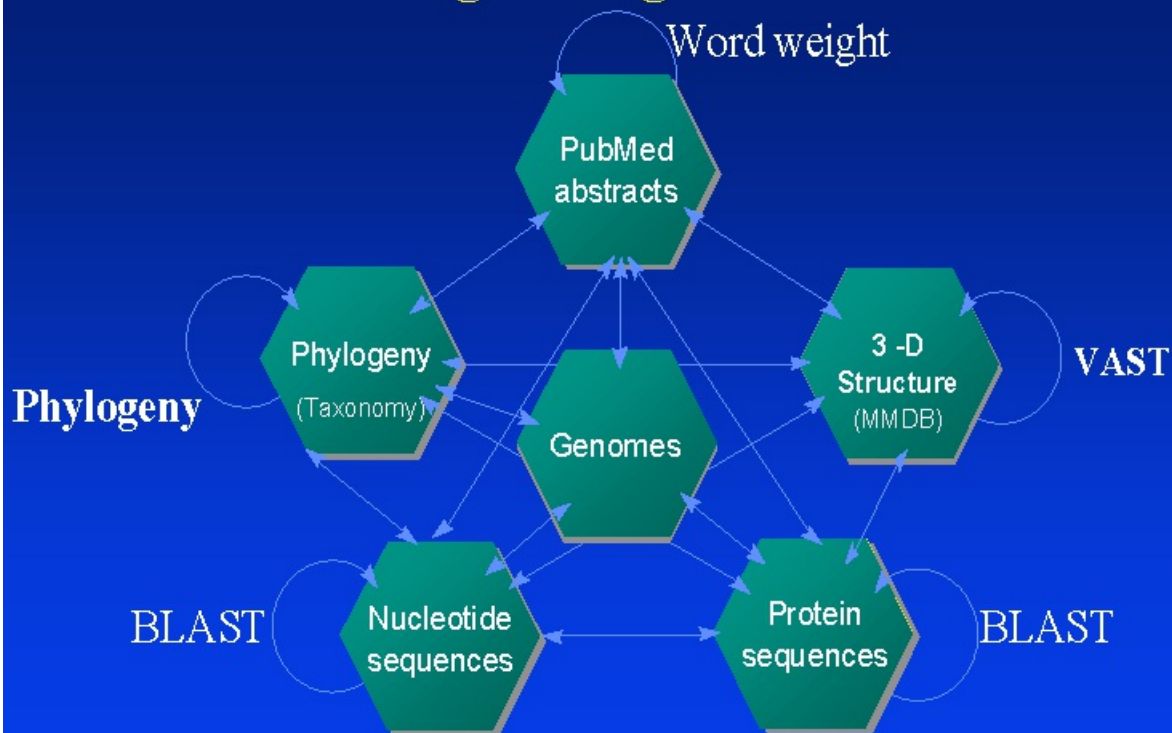
What is Entrez?

Offers integrated access to multiple databases

- Nucleotide sequences
 - GenBank
 - EMBL
 - RefSeq
 - PDB
- Protein sequences
 - SWISSPROT
 - GenPept
- Genomes, 3D structures, Pubmed entries and more!

Searchable by database, organism, sequence lengths, feature keys, molecular weights, gene locations, etc

Entrez: Neighboring and Hard Links



There may be useful data for the final project in these repositories

Accessing data via Entrez from Python

Open the `EntrezExample.ipynb` Jupyter Notebook file from [here](#)

This notebook will show you how to search and download from Entrez via Python

Questions:

- What is the downloaded sequence?
- What species is the downloaded sequence from?
- What does feature class of the sequence object return?
 - [The documentation may be helpful](#)

Write a new python function that takes two arguments, the first is an Entrez ID and the second is a name for a text file to output. The function should access the Entrez ID specified in the input and write the features to a text file, specified in the second argument.

If you aren't feeling comfortable working in the command line, the [Codecademy command line unit](#) is recommended