Python Semester Project Proposal

Daina Bouquin

Harvard-Smithsonian Center for Astrophysics (CFA)

November 19, 2015

The Goal: Create an interface that allows users to easily compare the average number of authors per paper among publications written on a variety of topics by researchers at the Harvard-Smithsonian Center for Astrophysics.

Description: I will create a simple interface to allow a user to enter (as raw input) three strings that will be used to populate an API query to the NASA Astrophysical Data System [1] using "A Python Module To Interact With NASA's ADS That Doesn't Suck" [2]. The resulting query will search the ADS for papers affiliated with the Harvard-Smithsonian Center for Astrophysics (CfA) bibliography that contain the raw input either as keywords or within the title of the paper. The results of the queries will be limited to publications authored by researchers at the CfA as I am interested in examining the publishing behaviors of researchers in my own community at the Center and to allow for simple prototyping of a tool that could be scaled up to include larger search results later. The results will also be limited to the Python module's default call to the ADS API (three pages) to prevent deep pagination and maxing out my daily API limit while testing the scripts. The results will subsequently be displayed as simple box and violin plots to show the different authorship behavior of researchers in different topics.

Modules to be used (*It is possible I won't use all modules on this list):

- import os
- import Tkinter*
- import tkFileDialog*
- import ads
- import numpy as np
- import scipy
- import pandas as pd*

I created **a simple prototype** of the proposed Python analysis using R which can be seen here [3] on RPubs.

[1] https://ui.adsabs.harvard.edu/
[2] http://https://github.com/andycasey/ads
[3] http://rpubs.com/dbouquin/ProjectTest