

# Lisc Python Package

Katie Brown

August 4, 2020

## 1 Introduction

Lisc collects and analyzes data from the National Center for Biotechnology Information (NCBI) database. This data can be counts and co-occurrences of search terms in the literature, text data and meta-data from articles that contain the search terms and can be used to collect citation data from DOIs. This data can be managed through data objects, saved, loaded and plotted.

## 2 Data

Using this package you can collect the counts data for the word co-occurrence data between terms as seen in Figure 1. Co-occurrence is defined as appearing together in different situations. We can also look at how many articles contain each of the specified search terms Figure 2. This figure shows the temporal lobe being the most studied part of the the brain. Another example of this packages function is it's ability to change NCBI databases, the above figures used the Pubmed database and Figure 3 uses the Nucleotide database. Nucleotide database hold genetic information about every studies species/organism on the planet. The example searched for five organisms and returned the number of articles that contained the search terms. Homo sapiens have the highest number of articles written.

## 3 Conclusions

Lisc is a great tool to collect data on articles using several databases. It can plot the data in various formats to effectively communicate the research

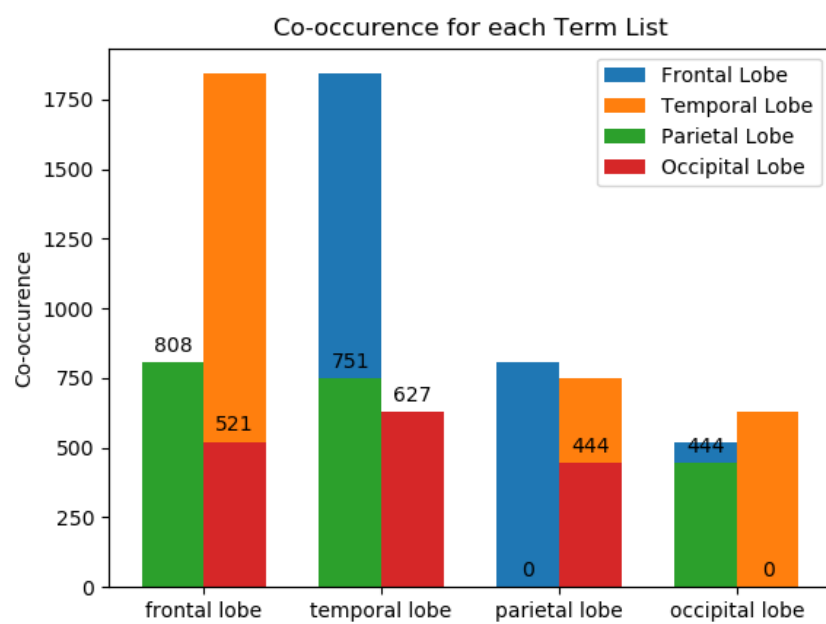


Figure 1: Matplotlib bar graph of the number of counts for word co-occurrence in the search term list.

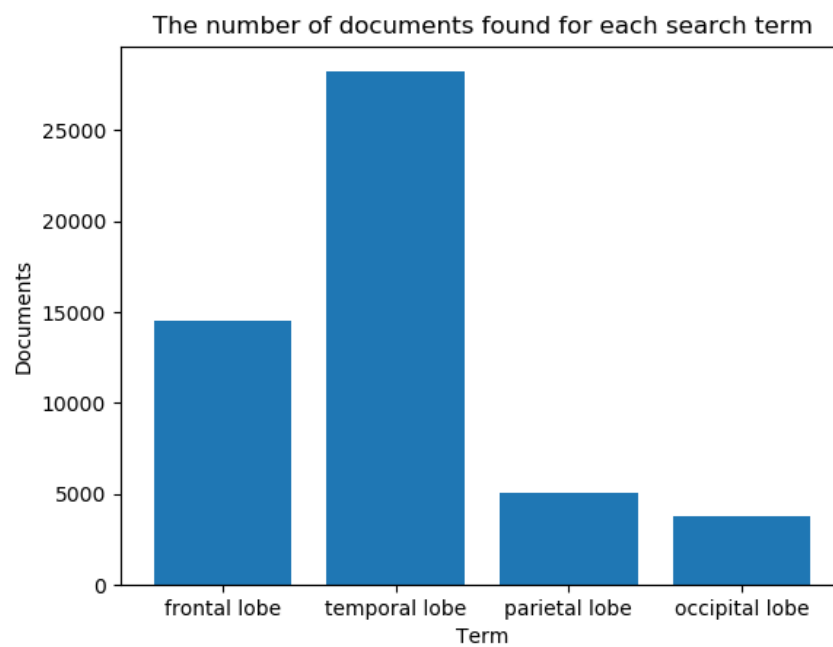


Figure 2: Matplotlib bar graph of the number of documents found for each search term.

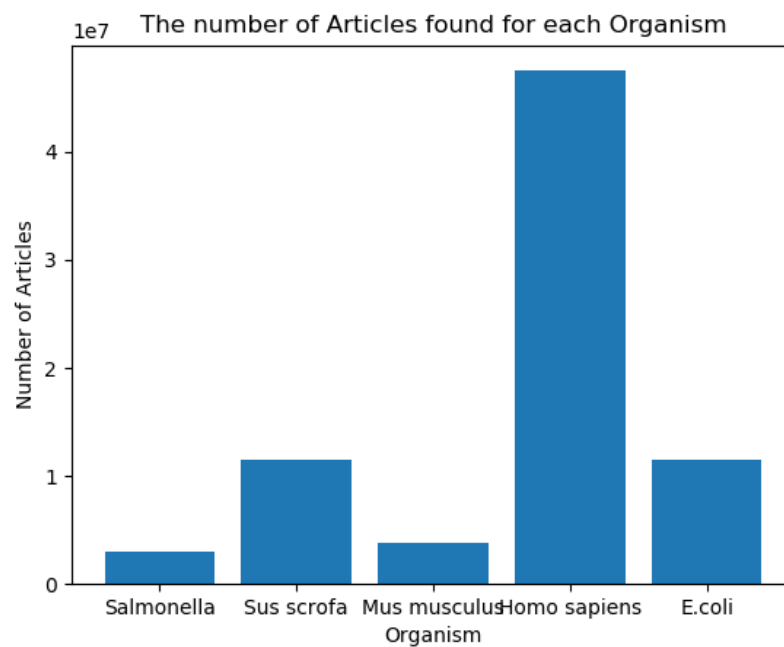


Figure 3: Matplotlib graph of the number of articles that contain "Salmonella", "E.coli", "Sus scrofa", "Homo sapiens", and "Mus musculus".

completed.

## 4 Reference

Donoghue, T. (2018) LISC: A Python Package for Scientific Literature Collection and Analysis. *Journal of Open Source Software*, 4(41), 1674. DOI: 10.21105/joss.01674