

rplos tutorial

The `rplos` package interacts with the API services of [PLOS](#) (Public Library of Science) Journals. In order to use `rplos`, you need to obtain [your own key](#) to their API services. Instruction for obtaining and installing keys so they load automatically when you launch R are on our GitHub Wiki page [Installation and use of API keys](#).

This tutorial will go through three use cases to demonstrate the kinds of things possible in `rplos`.

- Search across PLoS papers in various sections of papers
- Search for terms and visualize results as a histogram OR as a plot through time
- Text mining of scientific literature

Load package from CRAN

```
install.packages("rplos")
```

```
library(rplos)
```

Search across PLoS papers in various sections of papers

`searchplos` is a general search, and in this case searches for the term **Helianthus** and returns the DOI's of matching papers

```
searchplos(q = "Helianthus", fl = "id", limit = 5)
```

```
##              id
## 1 10.1371/journal.pone.0057533
## 2 10.1371/journal.pone.0045899
## 3 10.1371/journal.pone.0037191
## 4 10.1371/journal.pone.0051360
## 5 10.1371/journal.pone.0070347
```

Get only full article DOIs

```
searchplos(q = "*:~", fl = "id", fq = "doc_type:full", start = 0, limit = 5)
```

```
##              id
## 1 10.1371/journal.pntd.0001525
## 2 10.1371/journal.pone.0049273
## 3 10.1371/journal.pone.0031364
## 4 10.1371/journal.pone.0005841
## 5 10.1371/journal.pone.0005838
```

Get DOIs for only PLoS One articles

```
searchplos(q = "*:~", fl = "id", fq = "cross_published_journal_key:PLoSONE",
  start = 0, limit = 5)
```

```
##                                     id
## 1      10.1371/journal.pone.0049274/title
## 2      10.1371/journal.pone.0049274/abstract
## 3      10.1371/journal.pone.0049274/references
## 4      10.1371/journal.pone.0049274/body
## 5      10.1371/journal.pone.0049274/introduction
```

Get DOIs for full article in PLoS One

```
searchplos(q = ":*:*", fl = "id", fq = list("cross_published_journal_key:PLoSONE",
      "doc_type:full"), start = 0, limit = 5)
```

```
##                                     id
## 1 10.1371/journal.pone.0049273
## 2 10.1371/journal.pone.0031364
## 3 10.1371/journal.pone.0005841
## 4 10.1371/journal.pone.0005838
## 5 10.1371/journal.pone.0074814
```

Search for many terms

```
q <- c("ecology", "evolution", "science")
lapply(q, function(x) searchplos(x, limit = 2))
```

```
## [[1]]
##                                     id
## 1 10.1371/journal.pone.0059813
## 2 10.1371/journal.pone.0001248
##
## [[2]]
##                                     id
## 1 10.1371/journal.pbio.0050030
## 2 10.1371/journal.pbio.0030245
##
## [[3]]
##                                     id
## 1 10.1371/journal.pbio.0020122
## 2 10.1371/journal.pbio.1001166
```

Search on specific sections

A suite of functions were created as light wrappers around `searchplos` as a shorthand to search specific sections of a paper.

- `plosauthor` searches in authors
- `plosabstract` searches in abstracts
- `plostitle` searches in titles
- `plosfigtabcaps` searches in figure and table captions
- `plossubject` searches in subject areas

plosauthor searches across authors, and in this case returns the authors of the matching papers. the fl parameter determines what is returned

```
plosauthor(q = "Eisen", fl = "author", limit = 5)
```

```
##          author
## 1 Jonathan A Eisen
## 2 Jonathan A Eisen
## 3 Jonathan A Eisen
## 4 Jonathan A Eisen
## 5 Jonathan A Eisen
```

plosabstract searches across abstracts, and in this case returns the id and title of the matching papers

```
plosabstract(q = "drosophila", fl = "id,title", limit = 5)
```

```
##          id
## 1 10.1371/journal.pbio.0040198
## 2 10.1371/journal.pbio.0030246
## 3 10.1371/journal.pone.0012421
## 4 10.1371/journal.pbio.1000318
## 5 10.1371/journal.pbio.0030389
##
##          title
## 1 All for All
## 2 School Students as Drosophila Experimenters
## 3 Host Range and Specificity of the Drosophila C Virus
## 4 Genomic Responses to Abnormal Gene Dosage: The X Chromosome Improved on a Common Strategy
## 5 New Environments Set the Stage for Changing Tastes in Mates
```

plostitle searches across titles, and in this case returns the title and journal of the matching papers

```
plostitle(q = "drosophila", fl = "title,journal", limit = 5)
```

```
##          journal
## 1 PLoS Computational Biology
## 2 PLoS Biology
## 3 PLoS Genetics
## 4 PLoS ONE
## 5 PLoS Biology
##
##          title
## 1 Parametric Alignment of Drosophila Genomes
## 2 School Students as Drosophila Experimenters
## 3 Phenotypic Plasticity of the Drosophila Transcriptome
## 4 A Tripartite Synapse Model in Drosophila
## 5 Expression in Aneuploid Drosophila S2 Cells
```

Faceted search

Facet by journal

```
facetplos(q = "*:~", facet.field = "journal")
```

```
## $facet_queries
## NULL
##
## $facet_fields
## $facet_fields$journal
##
##      X1      X2
## 1      plos one 704650
## 2      plos genetics 34218
## 3      plos pathogens 29964
## 4      plos computational biology 25417
## 5      plos biology 24257
## 6      plos neglected tropical diseases 19520
## 7      plos medicine 17232
## 8      plos clinical trials 521
## 9      plos medicin 9
## 10     plos collections 5
##
##
## $facet_dates
## NULL
##
## $facet_ranges
## NULL
```

Using `facet.query` to get counts

```
facetplos(q = "*:~", facet.field = "journal", facet.query = "cell,bird")
```

```
## $facet_queries
##   term value
## 1 cell 83275
## 2 bird 8353
##
## $facet_fields
## $facet_fields$journal
##
##      X1      X2
## 1      plos one 704650
## 2      plos genetics 34218
## 3      plos pathogens 29964
## 4      plos computational biology 25417
## 5      plos biology 24257
## 6      plos neglected tropical diseases 19520
## 7      plos medicine 17232
## 8      plos clinical trials 521
## 9      plos medicin 9
## 10     plos collections 5
##
##
## $facet_dates
## NULL
```

```
##
## $facet_ranges
## NULL
```

Date faceting

```
facetplos(q = ".*", url = url, facet.date = "publication_date", facet.date.start = "NOW/DAY-5DAYS",
  facet.date.end = "NOW", facet.date.gap = "+1DAY")
```

```
## $facet_queries
## NULL
##
## $facet_fields
## NULL
##
## $facet_dates
## $facet_dates$publication_date
##      date value
## 1 2014-02-12T00:00:00Z 2508
## 2 2014-02-13T00:00:00Z 2334
## 3 2014-02-14T00:00:00Z  968
## 4 2014-02-15T00:00:00Z    0
## 5 2014-02-16T00:00:00Z    0
## 6 2014-02-17T00:00:00Z    0
##
##
## $facet_ranges
## NULL
```

Highlighted search

Search for the term *alcohol* in the abstracts of articles, return only 10 results

```
highplos(q = "alcohol", hl.fl = "abstract", rows = 2)
```

```
## $`10.1371/journal.pmed.0040151`
## $`10.1371/journal.pmed.0040151`$abstract
## [1] "Background: <em>Alcohol</em> consumption causes an estimated 4% of the global disease burden, p
##
##
## $`10.1371/journal.pone.0027752`
## $`10.1371/journal.pone.0027752`$abstract
## [1] "Background: The negative influences of <em>alcohol</em> on TB management with regard to delays :
```

Search for the term *alcohol* in the abstracts of articles, and return fragment size of 20 characters, return only 5 results

```
highplos(q = "alcohol", hl.fl = "abstract", hl.fragsize = 20, rows = 2)
```

```
## $`10.1371/journal.pmed.0040151`
## $`10.1371/journal.pmed.0040151`$abstract
```

```
## [1] "Background: <em>Alcohol</em>"
##
##
## $`10.1371/journal.pone.0027752`
## $`10.1371/journal.pone.0027752`$abstract
## [1] " of <em>alcohol</em> on TB management"
```

Search for the term *experiment* across all sections of an article, return id (DOI) and title fl only, search in full articles only (via `fq='doc_type:full'`), and return only 10 results

```
highplos(q = "everything:\"experiment\"", fl = "id,title", fq = "doc_type:full",
         rows = 2)
```

```
## $`10.1371/journal.pone.0039681`
## $`10.1371/journal.pone.0039681`$everything
## [1] " Selection of Transcriptomics <em>Experiments</em> Improves Guilt-by-Association Analyses Trans
##
##
## $`10.1371/journal.pone.0051016`
## $`10.1371/journal.pone.0051016`$everything
## [1] " Evolutionary Biology Breeding <em>Experience</em> Might Be a Major Determinant of Breeding
```

Search for terms and visualize results as a histogram OR as a plot through time

`plosword` allows you to search for 1 to K words and visualize the results as a histogram, comparing number of matching papers for each word

```
out <- plosword(list("monkey", "Helianthus", "sunflower", "protein", "whale"),
               vis = "TRUE")
out$table
```

```
##   No_Articles      Term
## 1         6894    monkey
## 2          230 Helianthus
## 3          607  sunflower
## 4         73707   protein
## 5          797     whale
```

```
out$plot
```

You can also pass in curl options, in this case get verbose information on the curl call.

```
plosword("Helianthus", callopts = list(verbose = TRUE))
```

```
## Number of articles with search term
##                               230
```

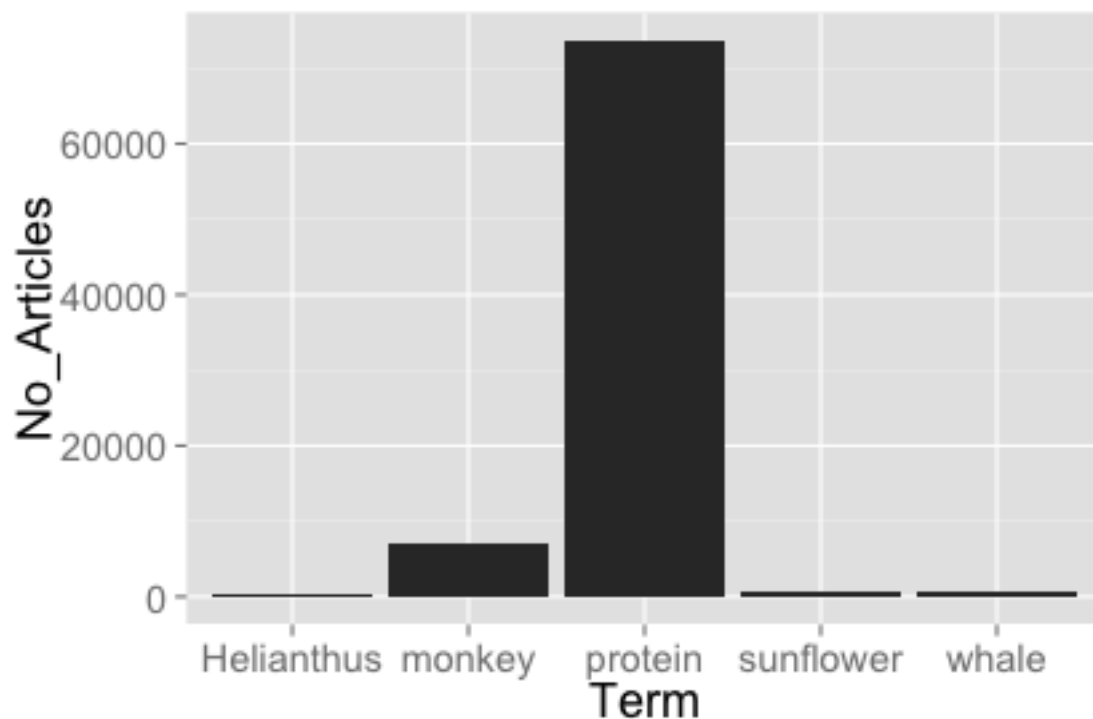


Figure 1: plot of chunk plosword1plot

Visualize terms

`plot_throughtime` allows you to search for up to 2 words and visualize the results as a line plot through time, comparing number of articles matching through time. Visualize with the `ggplot2` package, only up to two terms for now.

```
plot_throughtime(terms = "phylogeny", limit = 200)
```

OR using google visualizations through the `googleVis` package, check it your self using, e.g. (not shown here)

```
plot_throughtime(terms = list("drosophila", "flower"), limit = 200, gvis = TRUE)
```

...And a google visualization will render on your local browser and you can play with three types of plots (point, histogram, line), all through time. The plot is not shown here, but try it out for yourself!!

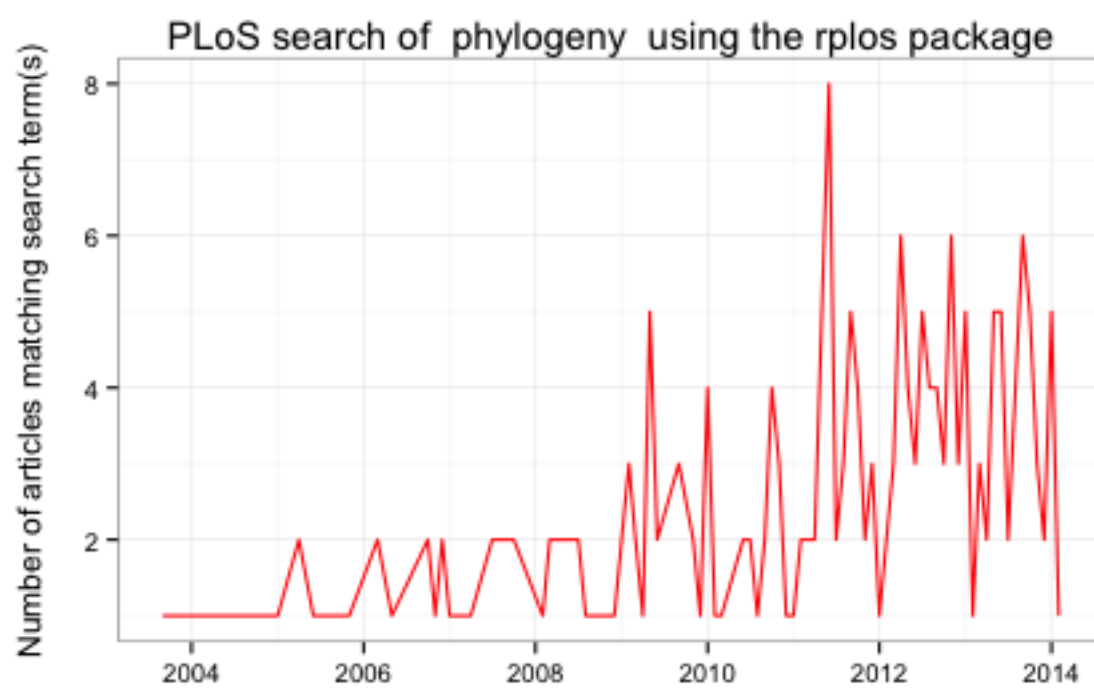


Figure 2: plot of chunk through time