

rOpenSci tools for textmining open source science literature

Scott Chamberlain^{*,a}

^a*rOpenSci, Museum of Paleontology, University of California, Berkeley, CA, USA*

Abstract

Corresponding Author:

Scott Chamberlain

rOpenSci, Museum of Paleontology, University of California, Berkeley, CA, USA

Email address: scott@ropensci.org

*Corresponding author

Email address: [scott\(at\)ropensci.org](mailto:scott(at)ropensci.org) (Scott Chamberlain)

9 Background. xxxx.

10 Methods. xxxx.

11 Results. xxxx.

Discussion. xxxx.

12 **Introduction**

13 Explosion of digital context online. But we need tools to facilitate textmining (i.e., rOpenSci)

14 Why text mine it.

15 How to text mine it.

16 Here, we introduce and give examples for textmining in R using rOpenSci packages, and related
17 packages.

18 **Text mining sources**

19 [Describe open source data for text mining that we link to in R, and those we don't link to in R.]

20 There is increasing open source scientific literature content available online. However, only a small
21 proportion of scientific journals provide access to their full content; whereas, many publishers provide
22 open access to their metadata only (Table 1).

Data Provider	Content Type	rOpenSci?	API
			Documentation
Public Library of Science (PLOS)	Full text/altmetrics	rplos	http://api.plos.org/
Springer	Full text on OA content	rspringer	http://dev.springer.com/
Pensoft	Full text/altmetrics	rpensoft	http://bit.ly/KYP0Zi
Nature Publishing Group	Metadata only	No	http://developers.nature.com/
Mendeley	Metadata only	rmendeley	http://dev.mendeley.com/
DataCite	Metadata only	rdatacite	http://oai.datacite.org/
Biodiversity Heritage Library	Full content	rbhl	http://bit.ly/KYQ1Rd

Data Provider	Content Type	rOpenSci?	API
			Documentation
Scopus (Elsevier)	Metadata only ????	No	http://bit.ly/J9S616

23 *Table 1. Sources of scientific literature, their content type provided via web services, whether rOpenSci*
24 *has an R packages for the service, and where to find the API documentation..*

Tool	Platform	Services	URL
PMC Miner	Python	PubMed	http://bit.ly/L29ekY
rOpenSci	R	See Table 1	http://ropensci.org/
PMC Miner	Python	PubMed	http://bit.ly/L29ekY
PMC Miner	Python	PubMed	http://bit.ly/L29ekY
PMC Miner	Python	PubMed	http://bit.ly/L29ekY
PMC Miner	Python	PubMed	http://bit.ly/L29ekY
PMC Miner	Python	PubMed	http://bit.ly/L29ekY
PMC Miner	Python	PubMed	http://bit.ly/L29ekY
PMC Miner	Python	PubMed	http://bit.ly/L29ekY
PMC Miner	Python	PubMed	http://bit.ly/L29ekY
PMC Miner	Python	PubMed	http://bit.ly/L29ekY
PMC Miner	Python	PubMed	http://bit.ly/L29ekY

25 *Table 2. Tools for text mining science literature..*

26 **How to text mine from R: Three case studies**

27 *Case study 1*

28 *Case study 2*

29 *Case study 3*

30 **Conclusions and future directions**

31 **Acknowledgments**

32 **Data Accessibility**

33 All scripts and data used in this paper can be found in the permanent data archive Zenodo under
34 the digital object identifier (DOI). This DOI corresponds to a snapshot of the GitHub repository at
35 <https://github.com/ropensci/textmine>. Software can be found at <https://github.com/ropensci/xxx>,
36 xxxx, all under MIT licenses.

37 **References**