taxize - taxonomic search and retrieval in R

Scott Chamberlain[1,*] and Eduard Szöcs[2,†]

[1]*Biology Department, Simon Fraser University, Canada.*
[2]*University Koblenz-Landau, Germany*

Keywords: taxonomy; R; sotware; data; API

---

[*] E-mail: myrmecocystus@gmail.com

[†] E-mail: szoe8822@uni-landau.de

I.   INTRODUCTION

some text here

II.   THE CASE FOR TAXIZE

There are a large suite of applications developed around the problem of searching for, resolving, and getting higher taxonomy for species names. For example, Linnaeus http://linnaeus.sourceforge.net/ provides ability to search for taxonomic names in documents and normalize names. In addition, there are many web interfaces to search for and normalize names such as Encyclopedia of Life's Global Names Resolver http://resolver.globalnames.org/, uBio tools http://www.ubio.org/index.php?pagename=sample_tools, and iPlant's Taxonomic Name Resolution Service http://tnrs.iplantcollaborative.org/.

All of these tools provide great ways to search for taxonomic names and resolve them in some cases. However, scientists ideally need a tool that can be used programmatically, and thus be made reproducible, and highly customizeable. The goal of taxize is to make it easy to create reproducible and easy to use workflows for searching for taxonomic names, resolving them, getting higher taxonomic names, and other tasks related to research dealing with species.

III.   DATA SOURCES

taxize uses many data sources, and more can easily be added.

TABLE I. Data sources used in taxize

Source name	Name search	Name resolution	Phylogeny	URL
Encyclopedia of Life	Yes	See GNR below	No	http://eol.org/
Integrated Taxonomic Resolution Service	Yes	Synonyms	No	X
iPlant Taxonomic Name Resolution Service	Yes	Yes	No	X
Phylomatic	No	X	No	X
uBio	Yes	X	No	X
Global Names Resolver (EOL)	Yes	X	No	X

IV.   USE CASES

There are a variety of use cases for which taxize is ideally suited, and few side cases in which taxize can be useful. We discuss five ideal use cases for taxize at length, and highlight the side cases in brief.

A.   Resolve taxonomic names

This is a common task in biology. We often have a list of species names and we want to know if a) we have the most up to date names, b) our names are spelled correctly, and c) if we have common names, we likely need the scientific names.

¡¡resolvenames, eval=TRUE, dev='png'¿¿= rnorm(10) @

B.   XX

C.   Retrieve higher taxonomic names

txt

D. Retrieve a phylogeny

text

E. Use ITIS locally for faster searches

text

F. Use ITIS locally for faster searches

text

V. CONCLUSION

some text here

VI. FUNDING

VII. ACKNOWLEDGEMENTS

some text here