

---

# rtoot: Collecting and Analyzing Mastodon Data

Journal Title  
XX(X):1–5  
©The Author(s) 0000  
Reprints and permission:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/ToBeAssigned  
www.sagepub.com/

SAGE

David Schoch\*<sup>1</sup>, Chung-hong Chan<sup>1</sup>

## Abstract

An implementation of calls designed to collect and organize Mastodon data via its Application Program Interfaces (API).

## Keywords

Mastodon; Social Media; Mobile Platform; API;

## Solution: Access the Mastodon API

Mastodon is a free and open source software which allows to run a self-hosted microblogging service, similar to Twitter. Servers running Mastodon can interoperate, meaning that their users can communicate across different instances. Together, all Mastodon instances form a large decentralized federation of social networking sites. Mastodon was first released in late 2016 but has not attracted as much attention by users and researchers compared to centralized social media platforms such as Twitter. Previous studies on Mastodon (e.g. Zulli et al., 2020; La Cava et al., 2022) classify it as an “alternative social media” platform, in contrast to “corporate social media” platforms such as Twitter. Due to the open source nature, Mastodon is also famous for being the technology behind far-right social media platforms Gab and Donald Trump’s Truth Social. These far-right platforms have also captured certain academic attention (e.g. Zhou et al., 2019; Zannettou et al., 2018).

The Twitter takeover by Elon Musk in 2022, however, has changed this alternative—if not fringe—status of the technology and sparked a huge wave of new registrations for instances running Mastodon (Huang, 2022). Mastodon will become increasingly more relevant for communication scholars who study online behaviour and phenomena. It is

---

<sup>1</sup>GESIS Leibniz Institute for the Social Sciences

### Corresponding author:

David Schoch, GESIS Leibniz Institute for the Social Sciences.

Email: [david.schoch@gesis.org](mailto:david.schoch@gesis.org)

important to note that Mastodon also is a mobile platform. Official mobile apps for the platform are recently released, although third-party mobile apps such as Tusky (Android) and Amaroq (iOS) have long been available.

Data collection from instances running Mastodon can be facilitated through the official Mastodon API ([Mastodon developers, 2022](#))\*. The R package `rtoot`<sup>†</sup> provides functions to authenticate for and interact with this API.

## Functionality

Many endpoints of the Mastodon API do not require authentication and calls can be made anonymously. However, instances can decide to require authentication for some endpoints and user action endpoints, such as sending a post, require an access token. Setting up a token only needs to be done once and the function `auth_setup()` guides users to obtain such a token from a particular instance.

Most other functions in `rtoot` are organized into four groups: `get_instance_*`() functions for collecting data about an instance; `get_account_*`() functions for collecting data about a user account; `get_timeline_*`() functions for collecting timelines (a series of statuses); and `stream_timeline_*`() functions for accessing the streaming API. Some exceptions are `search_accounts()` and `post_toot()`.

For instance, the following obtains a user access token from the “emacs.ch” instance. One can also obtain a public access token, which can only access public information.

```
library(rtoot)
auth_setup(instance = "emacs.ch", type = "user")
```

The following obtains information about the peers of the instance (other instances that this particular instance is aware of).

```
get_instance_peers(instance = "emacs.ch")
```

The activity of an instance can be obtained via

```
get_instance_activity(instance = "emacs.ch")
```

All `get_account_*`() functions require a user id instead of the user name as input. The only way to obtain the user id is to search for the user name. We have obtained the informed consent from the account owner of “chainsawriot” to collect data from his account.

---

\*Due to the decentralized nature of Mastodon, each instance has its own terms of service (ToS). Please check those terms to see whether this kind of cross-instance data collection is allowed. For instance, despite “scholar.social” allows API access, its ToS explicitly ban unapproved academic research. Also, you are not authorized to transfer these data to a third party, e.g. Google APIs. The research ethics is beyond the scope of this software presentation. Please consult [Zimmer and Kinder-Kurlanda \(2017\)](#) for general advice. Consult your ethics committees for specific.

<sup>†</sup>Before mid November 2022 (see <https://github.com/mastodon/mastodon/pull/18583>), a post on Mastodon was called a “toot”, similar to a “tweet” on twitter. The R package to interact with the Twitter API is called `rtweet` ([Kearney, 2019](#)), hence we chose the name `rtoot`.

```
search_accounts(query = "chainsawriot")
```

The following obtains the content published by that account.

```
get_account_statuses(id = "109337011845249544",
                     instance = "emacs.ch")
```

The tibble returned from `get_account_statuses` and any `get_timeline_*` and `stream_timeline_*` function, contains a list-column for the application used to post the status. To extract the name of the application, the following function can be used.

```
extract_application <- function(application) {
  if (length(application) == 0) {
    return(NA)
  } else {
    return(application$name)
  }
}
```

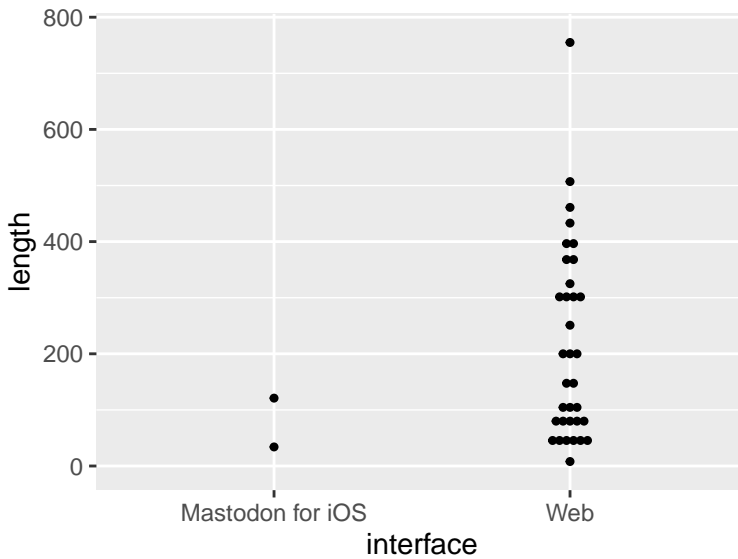
This auxiliary functions allows to filter out statuses that were posted via a mobile device or directly compare mobile with web-based posts. For example, one can visualize the length of toots from a mobile device (iOS) and the web interface (Figure 1).

```
library(tidyverse)
x <- get_account_statuses(id = "109337011845249544",
                         instance = "emacs.ch", limit = 30)
x %>% mutate(interface = map_chr(application,
                                extract_application),
             length = nchar(content)) %>%
  filter(!is.na(interface)) %>%
  ggplot(aes(x = interface, y = length)) +
  geom_dotplot(binaxis = "y", stackdir = "center",
              dotsize = 0.5)
```

More information can be found in the github wiki (<https://github.com/schochastics/rtoot/wiki/application>).

## Availability of the software and tutorials

`rtoot` is available on CRAN (<https://cran.r-project.org/web/packages/rtoot/index.html>) and from the official Github repository (<https://github.com/schochastics/rtoot>). The package contains tutorials on authentication, streaming, and a general introduction as vignettes. Additional tips on using the package are available from the official wiki (<https://github.com/schochastics/rtoot/wiki>) and the online documentation (<https://schochastics.github.io/rtoot/>).



**Figure 1.** Length of toots from different interfaces

## References

- Huang K (2022) What is mastodon and why are people leaving twitter for it? URL <https://www.nytimes.com/2022/11/07/technology/mastodon-twitter-elon-musk.html>.
- Kearney MW (2019) rtweet: Collecting and analyzing twitter data. *Journal of Open Source Software* 4(42): 1829. DOI:10.21105/joss.01829. URL <https://joss.theoj.org/papers/10.21105/joss.01829>. R package version 0.7.0.
- La Cava L, Greco S and Tagarelli A (2022) Network Analysis of the Information Consumption-Production Dichotomy in Mastodon User Behaviors. *Proceedings of the International AAAI Conference on Web and Social Media* 16: 1378–1382. DOI:10.1609/icwsm.v16i1.19391. URL <http://dx.doi.org/10.1609/icwsm.v16i1.19391>.
- Mastodon developers (2022) *Mastodon API*. URL <https://docs.joinmastodon.org/api/>.
- Zannettou S, Bradlyn B, De Cristofaro E, Kwak H, Sirivianos M, Stringini G and Blackburn J (2018) What is gab. *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18* DOI:10.1145/3184558.3191531. URL <http://dx.doi.org/10.1145/3184558.3191531>.
- Zhou Y, Dredze M, Broniatowski DA and Adler WD (2019) Elites and foreign actors among the alt-right: The gab social media platform. *First Monday* DOI:10.5210/fm.v24i9.10062. URL <http://dx.doi.org/10.5210/fm.v24i9.10062>.

- Zimmer M and Kinder-Kurlanda K (2017) *Internet research ethics for the social age: New challenges, cases, and contexts*. Peter Lang International Academic Publishers.
- Zulli D, Liu M and Gehl R (2020) Rethinking the “social” in “social media”: Insights into topology, abstraction, and scale on the mastodon social network. *New Media & Society* 22(7): 1188–1205. DOI:10.1177/1461444820912533. URL <http://dx.doi.org/10.1177/1461444820912533>.