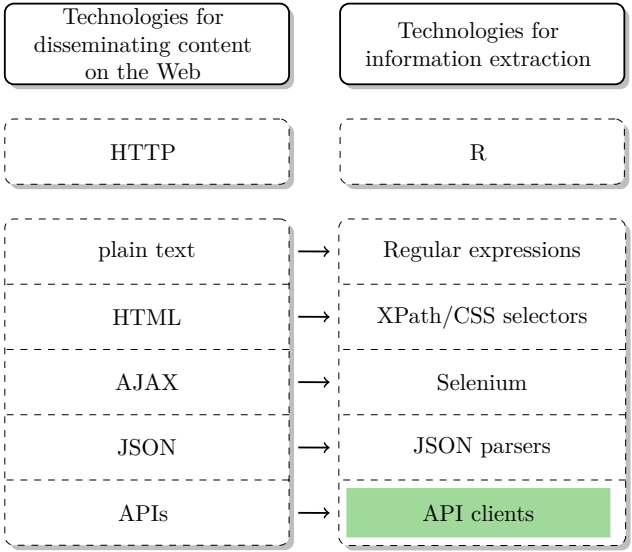
Introduction to Web Scraping with R

API Clients

Simon Munzert | IPSDS

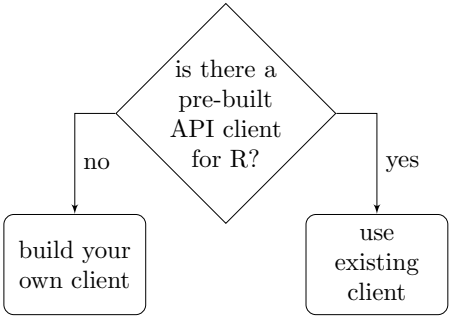
Technologies of the World Wide Web



Accessing APIs with R

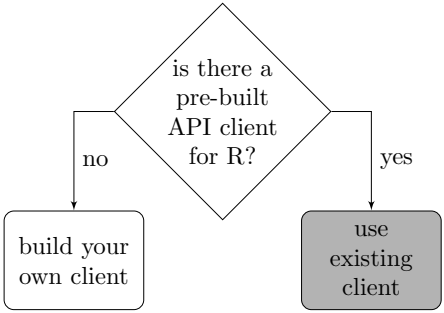
API access with R

There are basically **two scenarios:**



API access with R

There are basically **two scenarios:**



Here, we talk about the case where a client already exists.

API access with R

API clients

- provide interface to APIs
- hide API back-end
- let you stay in your programming environment

API access with R

API clients

- provide interface to APIs
- hide API back-end
- let you stay in your programming environment

Example

- the `rtweet` package provides an R client for Twitter
- lets you query data from the Twitter API with R commands
- you don't have to work with unfamiliar data formats (JSON) that is provided by the API—the client automatically transforms the incoming data into R objects

Finding API clients on the Web

General resources

List of APIs: http://www.programmableweb.com/apis

rOpenSci – collection of R API clients: https://github.com/ropensci/opendata

CRAN Task View: http://cran.r-project.org/web/views/WebTechnologies.html

Finding API clients on the Web

General resources

List of APIs: http://www.programmableweb.com/apis

rOpenSci – collection of R API clients: https://github.com/ropensci/opendata

CRAN Task View: http://cran.r-project.org/web/views/WebTechnologies.html

How to find the API client you need

- google "R package + name of website"
- search on the website for a "Developer" or "API" section (only if you don't find an R package that works)

Popular R API clients

package name	access to	more info
`rtweet`	Twitter Stream and REST API	http://rtweet.info/
`Rfacebook`	Facebook API	https://github.com/pablobarbera/Rfacebook
`ipapi`	ip-api.com's API	https://github.com/hrbrmstr/ipapi
`ggmap`	Google Maps/OpenStreetMap APIs	https://github.com/dkahle/ggmap
`eurostat`	Eurostat database	https://github.com/ropengov/eurostat
`rtimes`	New York Times APIs	https://cran.rstudio.com/web/packages/rtimes/index.html

Summary

Summary

- if the website you want to use as a data basis for your research provides access via an API, you can consider yourself lucky
- if there is a working R-based API client available, you hit the jackpot

Summary

- if the website you want to use as a data basis for your research provides access via an API, you can consider yourself lucky
- if there is a working R-based API client available, you hit the jackpot

Final considerations

- please always cite package authors when you use their work. run `citation("<package name>")` to see how the package should be cited
- sometimes API clients are not up to date. Consider to notify the author (e.g., by filing an issue on GitHub) or help update the package
- sometimes API clients provide functionality for only a fraction of the API's capability. It's always worth to check out the API documentation if you are looking for specific features