

Collecting Social Media Data

Two different methods:

1. Screen scraping: extract data from source code of website
2. Web APIs (application programming interface): use a set of structured https requests that return JSON or XML files

Collecting Social Media Data

Two different methods:

1. Screen scraping: extract data from source code of website
2. Web APIs (application programming interface): use a set of structured https requests that return JSON or XML files

Types of APIs:

1. RESTful APIs: queries for static information in current moment (e.g. user profiles, posts, etc.)
2. Streaming APIs: changes in users' data in real time (e.g. new messages, deletions, etc.)

Collecting Social Media Data

Two different methods:

1. Screen scraping: extract data from source code of website
2. Web APIs (application programming interface): use a set of structured https requests that return JSON or XML files

Types of APIs:

1. RESTful APIs: queries for static information in current moment (e.g. user profiles, posts, etc.)
2. Streaming APIs: changes in users' data in real time (e.g. new messages, deletions, etc.)

Rate limits

1. Restrictions on number of API calls by user and period of time
2. APIs are expensive!

Connecting with an API

Constructing a REST API call

- ▶ Baseline URL: `http://graph.facebook.com/`
- ▶ Parameters: `?ids=barackobama,johnmccain`

Connecting with an API

Constructing a REST API call

- ▶ Baseline URL: `http://graph.facebook.com/`
- ▶ Parameters: `?ids=barackobama,johnmccain`

Response often in JSON format. (example)

Connecting with an API

Constructing a REST API call

- ▶ Baseline URL: `http://graph.facebook.com/`
- ▶ Parameters: `?ids=barackobama,johnmccain`

Response often in JSON format. (example)

Authentication

- ▶ Most common is an open standard called OAuth
- ▶ Connections without sharing username and password, only temporary tokens that can be refreshed
- ▶ `httr` package in R implements most cases (examples)

Twitter and Facebook

R packages

- ▶ Twitter: `twitteR` for REST, `streamR` for Streaming
- ▶ Facebook: `Rfacebook`

Twitter and Facebook

R packages

- ▶ Twitter: `twitteR` for REST, `streamR` for Streaming
- ▶ Facebook: `Rfacebook`

Python: `tweepy` and `facebook-sdk`

Twitter and Facebook

R packages

- ▶ Twitter: `twitteR` for REST, `streamR` for Streaming
- ▶ Facebook: `Rfacebook`

Python: `tweepy` and `facebook-sdk`

Open-source code released by SMaPP lab (GitHUB)

Twitter and Facebook

R packages

- ▶ Twitter: `twitteR` for REST, `streamR` for Streaming
- ▶ Facebook: `Rfacebook`

Python: `tweepy` and `facebook-sdk`

Open-source code released by SMaPP lab (GitHUB)

Integration with `quanteda`