

Twitter Bot

Final Presentation

Sina Behdidka
Tim Mennicken
Robert Rose

University of Applied Sciences Cologne

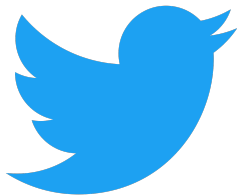
Tuesday the 4th of February, 2020

Table of Contents

- Introduction
- Methods
- Data Acquisition
 - Twitter API
 - GetOldTweets
- Pre-Processing
- Software
- Experiments
- Post-Processing
- References

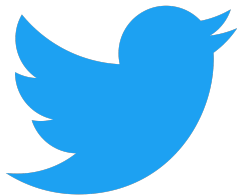
Data Acquisition

- ▶ Twitter developer platform
 - ▶ Twitter developer account needed
 - ▶ Tweepy: Python wrapper
 - ▶ Complete functionality

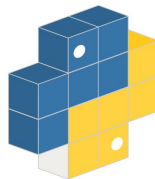


Data Acquisition

- ▶ Twitter developer platform
 - ▶ Twitter developer account needed
 - ▶ Tweepy: Python wrapper
 - ▶ Complete functionality



- ▶ GetOldTweets
 - ▶ Fetches Tweets from the website
 - ▶ Python package
 - ▶ Read only access



Data Acquisition

Twitter API

- ▶ Amount of Tweets limited
 - ▶ Request windows are separated in 15 minutes chunks
 - ▶ Specific amount of requests per window
 - ▶ Can be bypassed by cyclic requests paired with pauses

Data Acquisition

Twitter API

- ▶ Amount of Tweets limited
 - ▶ Request windows are separated in 15 minutes chunks
 - ▶ Specific amount of requests per window
 - ▶ Can be bypassed by cyclic requests paired with pauses
- ▶ Cannot go arbitrarily far to the past
 - ▶ Returns no Tweets older than roundabout a month
 - ▶ Cannot be bypassed

Data Acquisition

Twitter API

- ▶ Amount of Tweets limited
 - ▶ Request windows are separated in 15 minutes chunks
 - ▶ Specific amount of requests per window
 - ▶ Can be bypassed by cyclic requests paired with pauses
- ▶ Cannot go arbitrarily far to the past
 - ▶ Returns no Tweets older than roundabout a month
 - ▶ Cannot be bypassed
- ▶ Account got blacklisted
 - ▶ No further access to the Twitter API
 - ▶ Got unblocked on request

Data Acquisition

GetOldTweets

- ▶ No limitation
 - ▶ Arbitrary amount of Tweets
 - ▶ No restrictions with respect to publication date

Data Acquisition

GetOldTweets

- ▶ No limitation
 - ▶ Arbitrary amount of Tweets
 - ▶ No restrictions with respect to publication date
- ▶ Shortened functionality
 - ▶ Limited meta data of Tweets
 - ▶ No functionality for publishing Tweets

Data Acquisition

GetOldTweets

- ▶ No limitation
 - ▶ Arbitrary amount of Tweets
 - ▶ No restrictions with respect to publication date
- ▶ Shortened functionality
 - ▶ Limited meta data of Tweets
 - ▶ No functionality for publishing Tweets
- ▶ Better suited for getting large data sets

Pre-Processing

Particular Content

- ▶ Weblinks
- ▶ Picture & video links
- ▶ Punctuation symbols
- ▶ Special characters
- ▶ Retweets
- ▶ Hashtags
- ▶ Username references

Pre-Processing

Particular Content

Content we keep:

1. Selected punctuation symbols
point, comma, exclamation mark, interrogation mark, colon
and hash
2. Hashtags
3. References to usernames

Pre-Processing

Space characters

- ▶ Union of several space characters
- ▶ The Tokenizer splits the input text at space characters

Pre-Processing

Space characters

- ▶ Union of several space characters
- ▶ The Tokenizer splits the input text at space characters

Why we need to add spaces

house. \Rightarrow ["house."]

house . \Rightarrow ["house", "."]

Pre-Processing

Termination Symbol

Twitter Bot

Sina Behdidka, Tim Mennicken, Robert Rose
04.02.2020

Slide 8 of 11

Technology
Arts Sciences
TH Köln

Software

Structure

Twitter Bot

Sina Behdidka, Tim Mennicken, Robert Rose
04.02.2020

Slide 9 of 11

Technology
Arts Sciences
TH Köln

Software

Interaction

Twitter Bot

Sina Behdidka, Tim Mennicken, Robert Rose
04.02.2020

Slide 10 of 11

Technology
Arts Sciences
TH Köln

References

- ▶ Twitter Developers – <https://developer.twitter.com/en.html>
- ▶ GetOldTweets – <https://github.com/Jefferson-Henrique/GetOldTweets-python>