Why

- extract data from a report that is released weekly and published online
- grab the schedule for your favorite sports team as it's released
- find the release dates for upcoming movies in your favorite genre
- be notified automatically when a website changes
- obtain datasets to for training/reinforced learning/modelling
- text classification, sentiment analysis, data enrichment, projections, graphing or plotting

Libraries and Modules

- requests
- beautifulSoup
- pandas
- csv
- json
- datetime
- Scrapy/Selenium
- BeautifulSoup

Twitter Analysis

Twitter developer account + API

Top endpoints
- Below are a few examples of the Twitter API endpoints, requests, and responses. These examples use twurl—a command-line application that can be used to make authenticated requests to the Twitter platform. twurl is like curl, except that it abstracts away OAuth details once you configure it with your keys.
- Search API
- Ads API
- Engagement API
- Direct Message API
- Account Activity API
- Embed a Tweet

Python - Twitter Library

This library provides a pure Python interface for the Twitter API. It works with Python 3.X

Twitter provides a service that allows people to connect via the web, IM, and SMS. Twitter exposes a web services API and this library is intended to make it even easier for Python programmers to use.

Sentiment Analysis

Sentiment analysis (also known as opinion mining or emotion AI) refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine.

Software

There are many software tools available that can be used to customize web-scraping solutions. This software may attempt to automatically recognize the data structure of a page or provide a recording interface that removes the necessity to manually write web-scraping code, or some scripting functions that can be used to extract and transform content, and database interfaces that can store the scraped data in local databases. Some web scraping software can also be used to extract data from an API directly.

Data Ethics and Legal Issues

The legality of web scraping varies across the world. In general, web scraping may be against the terms of use of some websites, but the enforceability of these terms is unclear. There are also many concerns regarding cross-user sensitive personal information for users and consumer protection.

Resources

- https://developer.twitter.com/en
- https://tools atravex.com/tw
- https://www.nihilhosting.com/tools/user-agent/