



DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Lecture 2.1 - Collecting Data



The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

How to Collect Data?

Method

Effort

Download

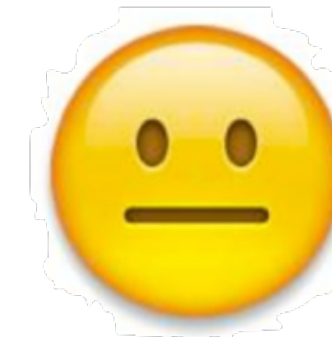
Low



API

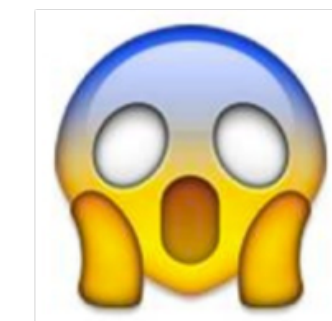
(Application program interface)

Medium



Scrape/Crawl

High












Data You Can Just Download

- NYC Taxi data: Trip (11GB), Fare (7.7GB)
- StackOverflow (xml)
- Wikipedia (data dump)
- Atlanta crime data (csv)
- Soccer statistics
- ...

Collect Data via APIs

- Google Data API (e.g., Google Maps Directions API)
<https://developers.google.com/gdata/docs/directory>
- Twitter (small subset)
<https://dev.twitter.com/streaming/overview>
- Last.fm (Pandora has unofficial API)
- Flickr
- data.nasa.gov
- data.gov
- Facebook (your friends only)

API	GData Status	See Also
 Google Analytics Data Export API	Replaced by Google Analytics Core Reporting API (starting at version 2.4).	Migration Guide: Moving from v2.3 APIs to v2.4 & v3.0
 Google Apps Provisioning API	Shut down. Replaced by the Admin SDK Directory API .	Current Google Apps APIs
 Google Base Data API	Not available since June 1, 2011. Replaced by the Content API for Shopping .	New Shopping APIs and Deprecation of the Base API
 Blogger Data API	Replaced by the latest Blogger API .	
 Google Book Search API	Shut down. Replaced by Google Books API Family .	Google books API searching by ISBN (on Stack Overflow)
 Google Calendar API v2	Shut down. Replaced by latest Google Calendar API .	
 Google Code Search Data API	Shut down in Jan 15, 2012. No replacement API.	A fall sweep (Google blog post)
 Google Contacts API	GData version is still live. Replaced by Google People API for read-only access.	Google Contacts API Google People API
 Google Documents List Data API	Shut down. Replaced by Google Drive API .	
 Google Finance Portfolio Data API	Shut down. No replacement API.	Spring cleaning for some of our APIs (Google blog post)
 Google Health Data API	The product was discontinued as of January 1, 2013. No replacement API.	An update on Google Health and Google PowerMeter

<https://developers.google.com/gdata/docs/directory>

Data Unavailable via APIs

- Amazon (reviews, product info)
- ESPN
- eBay
- Google Play
- Google Scholar
- ...



DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Thank You