**Follow instructions here to download Anaconda Python 3, complete with all the necessary modules.**

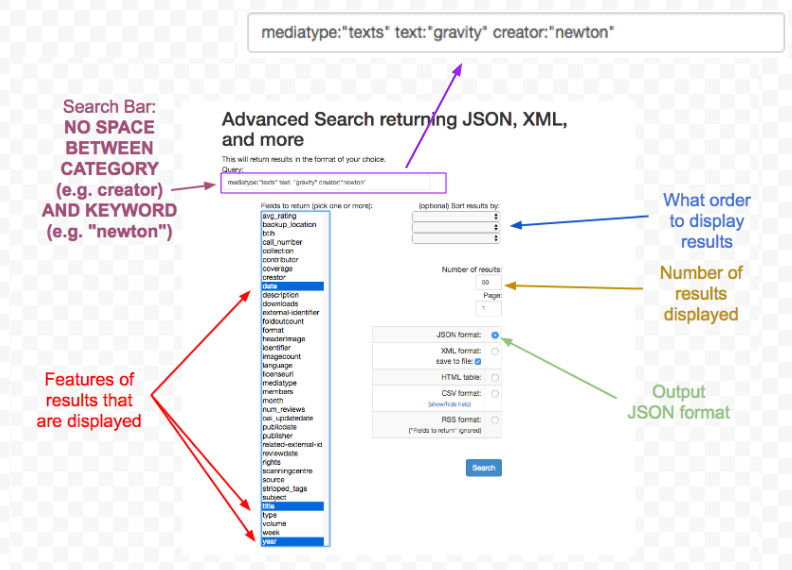
[Install Anaconda](https://www.cs.hmc.edu/twiki/bin/view/CS5/OwnMachines)

**Note:** I included print statements just to help visualize how the function is working. These can be commented out when you gather large amounts of data because they slow down the program.

**Archive.org**

The Internet Archive API allows you to quickly navigate through thousands of texts.

<https://archive.org/advancedsearch.php>

Go to "Advanced Search returning JSON, XML, and more"

You may sometimes need to use OR, AND (default for a space), or TO (for ranges) to broaden or specify a search. If no category is specified, all categories are checked for the keyword:





Complete details and resources [here](https://archive.org/advancedsearch.php) under "**Explanation of Terms, Operators, Queries, and Grouping"**

**Notes**

* old texts use "f" instead of "s" in some instances
* Most texts not indexed for Search Inside API (full text search)
* Select "identifier", "date", and "year" for return fields for functions in archive\_scraping.py to work properly

**New York Times**

**Get an API key from the New York Times**

[Get API key](https://developer.nytimes.com/signup)

* Input name, email, and for website, do <your\_name>.com

**In case you are interested, here are the full documentations of the resources being used in nytimes\_scraping.py.**

**Article Search (by query field)**

[Article Search README](https://developer.nytimes.com/article_search_v2.json#/README)

[Article Search Documentation](https://developer.nytimes.com/article_search_v2.json?q=obama&facet_field=source&begin_date=20120101&end_date=20121231&facet_filter=true#/Documentation/GET/articlesearch.json)

**Archive Search (by month and year)**

[Archive Search README](https://developer.nytimes.com/archive_api.json#/README)

[Archive Search Documentation](https://developer.nytimes.com/archive_api.json#/Documentation)

**Notes**

* Full-text search only for 1980-present (type\_of\_material: "news")
* You can return "snippet" and/or "headline" parameters to get a semblance of word counts for items since 1851-1923 and 1923-1980 (type\_of\_material: "articles")
* 1000 requests per day (10 articles per request)
* 1851–1922: Full articles are free to access for all visitors.
* 1923–1980: Full articles are accessible to paid subscribers, or are available for purchase.
* 1981–Present: Full articles are accessible by all visitors, but count toward the limited number of free articles per month.