

INLS 560
Programming for
Information Professionals

Accessing
Web Data



UNC
SCHOOL OF INFORMATION
AND LIBRARY SCIENCE

Joan Boone

jpboone@email.unc.edu

Part 1: Parsing Apple Feed JSON Data

Part 2: Accessing Web Data with URLs

Web Data

Web data is often 'open data' that is free and publicly available so that anyone can access and use it without restrictions.

W3C Recommendation: [Data on the Web Best Practices](#)

- As data becomes more ubiquitous, and data sets become larger and more complex, processing by computers becomes ever more crucial.
- Data becomes useful when it has been processed and transformed into information.

- [Best Practice for Data Formats](#)

Make data available in a machine-readable standardized data format that is easily parseable including but not limited to CSV, XML, HDF5, JSON and RDF serialization syntaxes like RDF/XML and JSON-LD

Many Web Data Sources...for example, Apple RSS Feed Generator

RSS Feed Generator

Feed URL

<https://rss.applemarketingtools.com/api/v2/us/music/most-played/10/albums.json>

Copy

Feed Settings: select type, storefront, chart, result limit, format

Feed URL specifies the JSON data file for your selection. Copy and paste the URL in your browser to see the data.

Feed Settings

Use the dropdowns to pick setting to generate a new RSS feed URL.

Media Type	Storefront	Type
Music	United States	Albums
Chart	Result Limit	Format
Most Played	10	JSON

Feed Preview displays top 10 most played music albums (United States) using JSON format

Feed Preview

Apple Music: Albums



One Thing At A Time
Morgan Wallen



CALL ME IF YOU GET LOST: The Estate Sale
Tyler, The Creator



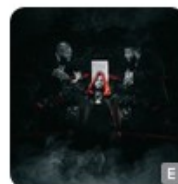
Timeless
Davido



PORTALS
Melanie Martinez



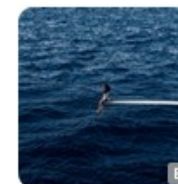
Gettin' Old
Luke Combs



I'M REALLY LIKE THAT
DJ Drama



The Great Escape
Larry June & The Alchemist



SOS
SZA



Her Loss
Drake & 21 Savage



MAÑANA SERÁ BONITO
KAROL G

Apple feed JSON Data for top 10 most played albums

```
{
  "feed": {
    . . .
    "results": [
      {
        "artistName": "Morgan Wallen",
        "id": "1667990565",
        "name": "One Thing At A Time",
        "releaseDate": "2023-03-03",
        "kind": "albums",
        "artistId": "829142092",
        "artistUrl": "https://music.apple.com/us/artist/morgan-wallen/829142092",
        "contentAdvisoryRating": "Explicit",
        "artworkUrl100": "https://is3-ssl.mzstatic.com/image/.../100x100bb.jpg",
        "genres": [
          {
            "genreId": "6",
            "name": "Country",
            "url": "https://itunes.apple.com/us/genre/id6"
          },
          {
            "genreId": "34",
            "name": "Music",
            "url": "https://itunes.apple.com/us/genre/id34"
          }
        ],
        "url": "https://music.apple.com/us/album/one-thing-at-a-time/1667990565"
      },
      . . .
    ]
  }
}
```

apple_music.json

Extracting selected data from the Apple feed

```
{
  "feed": {
    . . .
    "results": [
      {
        "artistName": "Morgan Wallen",
        "id": "1667990565",
        "name": "One Thing At A Time",
        "releaseDate": "2023-03-03",
        "kind": "albums",
        "artistId": "829142092",
        "artistUrl": "https://music.apple.com/us/artist/morgan-wallen/829142092",
        "contentAdvisoryRating": "Explicit",
        "artworkUrl100": "https://is3-ssl.mzstatic.com/image/.../100x100bb.jpg",
        . . .
      }
    ]
  }
}
```

```
Morgan Wallen
One Thing At A Time
2023-03-03
```

Output after parsing the JSON data and selecting artist name, album name, and release date

```
Tyler, The Creator
CALL ME IF YOU GET LOST: The Estate Sale
2021-06-25
```

```
Davido
Timeless
2023-03-31
```

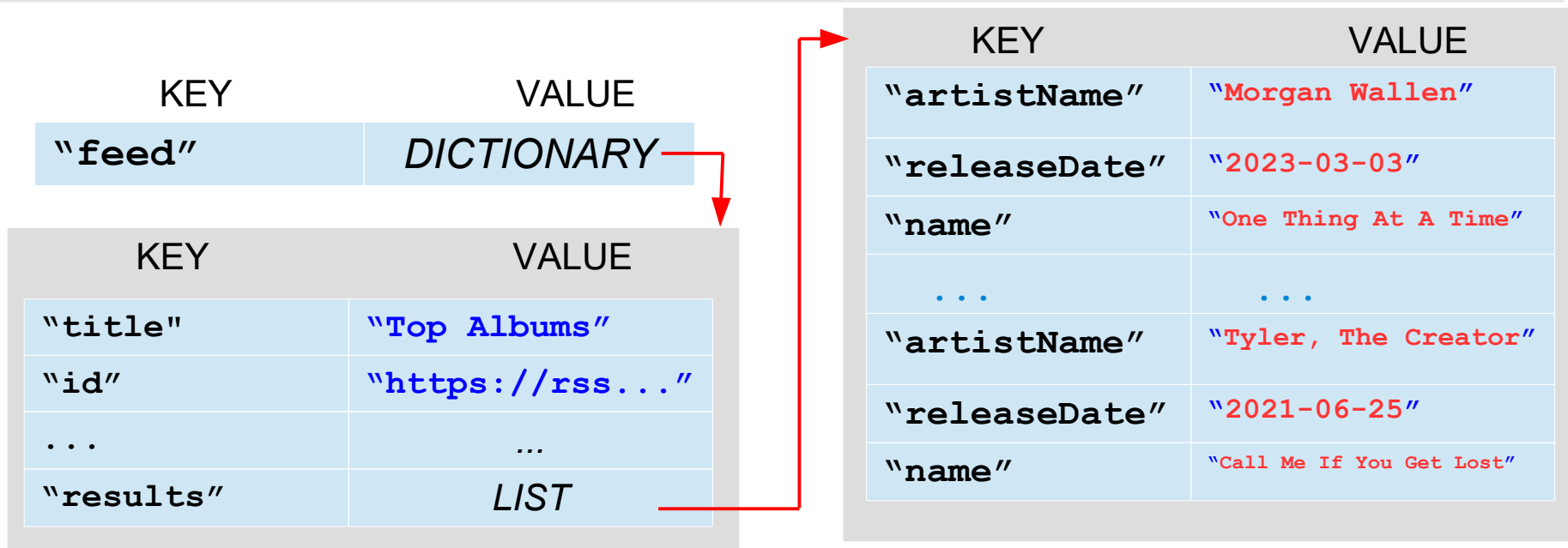
```
Melanie Martinez
PORTALS
2023-03-31
```

parse_apple_feed_from_file.py apple_music.json

Apple feed JSON Data for top 10 most played albums

```
{
  "feed": {
    . . .
    "results": [
      {
        "artistName": "Morgan Wallen",
        "id": "1667990565",
        "name": "One Thing At A Time",
        "releaseDate": "2023-03-03",
        "kind": "albums",
        "artistId": "829142092",
        "artistUrl": "https://music.apple.com/us/artist/morgan-wallen/829142092",
        "contentAdvisoryRating": "Explicit",
        "artworkUrl100": "https://is3-ssl.mzstatic.com/image/.../100x100bb.jpg",
        . . .
      }
    ]
  }
}
```

It is *very important* to understand the data structures defined by the JSON, so that you can parse it correctly!



Parsing Apple feed JSON data

```
import json
...
feed_dictionary = json.loads(feed_content)

feed_details = feed_dictionary['feed']

results_list = feed_details['results']

for result in results_list:
    artistName = result['artistName']
    name = result['name']
    date = result['releaseDate']

    print(artistName)
    print(name)
    print(date)
    print()
```

Import the module
for the JSON parser

`loads` parses the JSON text and
converts it to a Python dictionary

The dictionary has one entry where
key = `'feed'` and
value = a dictionary that describes
the details of the feed

The `feed_details` dictionary has
an entry with a
key = `'results'` and
value = list of results from the
feed

Each `result` is a dictionary with keys
for `'artistName'`,
`'name'`, and
`'releaseDate'`

Exercise: Update Apple Feed Example

- Extract and print the URL for the album
- Extract and print the genres for the album

Morgan Wallen

One Thing At A Time

2023-03-03

<https://music.apple.com/us/album/one-thing-at-a-time/1667990565>

Country Music

Tyler, The Creator

CALL ME IF YOU GET LOST: The Estate Sale

2021-06-25

<https://music.apple.com/us/album/call-me-if-you-get-lost-the-estate-sale/1679454273>

Music Hip-Hop/Rap

Davido

Timeless

2023-03-31

<https://music.apple.com/us/album/timeless/1675379314>

Music

Melanie Martinez

PORTALS

2023-03-31

<https://music.apple.com/us/album/portals/1673604325>

Alternative Music Pop

...

Output

Part 1: Parsing Apple Feed JSON Data

Part 2: Accessing Web Data with URLs

Accessing Web Data with URLs

To access web data, you need a URL (Uniform Resource Locator) that defines a unique address for a web resource. Python has a built-in module, `urllib`, that is used to fetch content associated with a URL.

- `urllib.request` for opening and reading
- `urllib.error` for handling exceptions

Steps

- Import the required modules

```
from urllib.request import urlopen
from urllib.error import URLError, HTTPError
```
- Open the URL and save the HTTP response to a variable

```
response = urlopen(feed_url)
```
- Read the data from the response and save to a variable

```
feed_content = response.read()
```
- Add exception handlers

```
HTTPError to handle problems with HTTP requests
URLError to handle general URL-related problems
```

Accessing Apple feed JSON data by URL

```
from urllib.request import urlopen
from urllib.error import URLError, HTTPError
import json

try:
    feed_url = "https://rss.applemarketingtools.com/api/v2/us/music/most-played/10/albums.json"
    response = urlopen(feed_url)      # open the URL and save HTTP response object
    feed_content = response.read()    # read the data from the response and save

    # Parse the json data, extract and output information of interest
    feed_dictionary = json.loads(feed_content)

    feed_details = feed_dictionary['feed']
    results_list = feed_details['results']
    for result in results_list:
        artistName = result['artistName']
        name = result['name']
        date = result['releaseDate']

        print(artistName, name, date, sep="\n")
        print()
except ValueError as err:
    print('An error occurred trying to decode the json text')
    print(err)
except HTTPError as err:
    print('Server could not fulfill the request.')
    print(err)
except URLError as err:
    print('Failed to reach a server.')
    print(err)
except Exception as err:
    print('An error occurred: ', err)
```

parse_apple_feed_from_url.py

Exercise: Modify Apple Feed URL

- Go to [Apple Feed Generator](#) and modify Feed Settings
- Copy and paste the generated Feed URL and substitute into the program to view the different results.
- NOTE: the JSON format is not the same for all Feed Settings. For example, the format is the same for music albums and music videos, but this format is different from music playlists.

<https://is3-ssl.mzstatic.com/image/thumb/Music123/.../100x100bb.jpg>

One Thing At A Time

2023-03-03

<https://music.apple.com/us/album/one-thing-at-a-time/1667990>

<https://is5-ssl.mzstatic.com/image/thumb/Music126/.../100x100bb.jpg>

CALL ME IF YOU GET LOST: The Estate Sale

2021-06-25

<https://music.apple.com/us/album/call-me-if-you-get-lost-the-estate-sale/1667990>

<https://is2-ssl.mzstatic.com/image/thumb/Music126/.../100x100bb.jpg>

PORTALS

2023-03-31

<https://music.apple.com/us/album/portals/1673604325>

<https://is5-ssl.mzstatic.com/image/thumb/Music116/.../100x100bb.jpg>

Timeless

2023-03-31

<https://music.apple.com/us/album/timeless/1675379314>

<https://is5-ssl.mzstatic.com/image/thumb/Music113/.../100x100bb.jpg>

Gettin' Old

2023-03-24

<https://music.apple.com/us/album/gettin-old/1666738524>

...

Re-purposing Content: Apple Feed JSON data to HTML

apple_feed_to_HTML.py, style.css

Apple Music: Top 10 Albums



One Thing At A Time
2023-03-03



CALL ME IF YOU GET
LOST: The Estate Sale
2021-06-25



PORTALS
2023-03-31



Timeless
2023-03-31



Gettin' Old
2023-03-24



I'M REALLY LIKE THAT
2023-03-31



The Great Escape
2023-03-31



SOS
2022-12-09



Her Loss
2022-11-04



MAÑANA SERÁ BONITO
2023-02-24

Output: apple_feed.html