PFCH Final Project

**No News is Old News**

To start with, let’s look at the metadata available to compare the NYTimes and The Guardian’s information.

Using The Guardian’s “‘all-fields’ : ‘all’” (see **TheGuardian\_all-metadata\_no-API-key.py**) reveals the following data structure and metadata:

{

"id": "tv-and-radio/2022/dec/30/from-wednesday-and-matilda-to-nye-fireworks-the-best-things-for-children-to-watch-this-weekend",

"type": "article",

"sectionId": "tv-and-radio",

"sectionName": "Television & radio",

"webPublicationDate": "2022-12-29T23:00:32Z",

"webTitle": "From Wednesday and Matilda to NYE fireworks: the best things for children to watch this weekend",

"webUrl": "https://www.theguardian.com/tv-and-radio/2022/dec/30/from-wednesday-and-matilda-to-nye-fireworks-the-best-things-for-children-to-watch-this-weekend",

"apiUrl": "https://content.guardianapis.com/tv-and-radio/2022/dec/30/from-wednesday-and-matilda-to-nye-fireworks-the-best-things-for-children-to-watch-this-weekend",

"fields": {

"headline": "From Wednesday and Matilda to NYE fireworks: the best things for children to watch this weekend",

"standfirst": "<p>Exhausted by the festive season? You\u2019re not alone. Here\u2019s a bunch of films and streams to keep kids of all ages occupied around Australia</p><ul><li><a href=\"https://www.theguardian.com/newsletters/2019/oct/18/saved-for-later-sign-up-for-guardian-australias-culture-and-lifestyle-email?CMP=cvau\_sfl\">Get our weekend culture and lifestyle email</a></li></ul>",

"trailText": "Exhausted by the festive season? You\u2019re not alone. Here\u2019s a bunch of films and streams to keep kids of all ages occupied around Australia",

"byline": "Luke Buckmaster with Nick Miller",

"main": "<figure class=\"element element-image\" data-media-id=\"f09c1900034e25aa51bea5da0b2fab08f1c38522\"> <img src=\"https://media.guim.co.uk/f09c1900034e25aa51bea5da0b2fab08f1c38522/0\_0\_2500\_1500/1000.jpg\" alt=\"Composite image of Puss in Boots, Jenna Ortega as Wednesday Addams and Alisha Weir as Matlida\" width=\"1000\" height=\"600\" class=\"gu-image\" /> <figcaption> <span class=\"element-image\_\_caption\">Our guide to on-screen summer offerings to keep children entertained.</span> <span class=\"element-image\_\_credit\">Composite: Universal Pictures, Alamy, Netflix, Dan Smith, Getty Images</span> </figcaption> </figure>",

"body": ,

"wordcount": "1090",

"firstPublicationDate": "2022-12-29T23:00:32Z",

"isInappropriateForSponsorship": "false",

"isPremoderated": "false",

"lastModified": "2023-01-05T06:38:47Z",

"productionOffice": "AUS",

"publication": "theguardian.com",

"shortUrl": "https://www.theguardian.com/p/n2cmc",

"shouldHideAdverts": "false",

"showInRelatedContent": "true",

"thumbnail": "https://media.guim.co.uk/f09c1900034e25aa51bea5da0b2fab08f1c38522/0\_0\_2500\_1500/500.jpg",

"legallySensitive": "false",

"lang": "en",

"isLive": "true",

"bodyText":

"charCount": "6647",

"shouldHideReaderRevenue": "false",

"showAffiliateLinks": "false",

"bylineHtml": "<a href=\"profile/luke-buckmaster\">Luke Buckmaster</a> with <a href=\"profile/nick-miller-\">Nick Miller</a>",

"showTableOfContents": "false"

},

"isHosted": false,

"pillarId": "pillar/arts",

"pillarName": "Arts"

},

**NYTimes (see NYtimes\_all-metadata.py)**

{

"status": "OK",

"copyright": "Copyright (c) 2023 The New York Times Company. All Rights Reserved.",

"response": {

"docs": [

{

"abstract": "Jordan wasn\u2019t vocal about social justice like today\u2019s N.B.A. stars, but his secret brand of activism is a key reason they have the spotlight now.",

"web\_url": "https://www.nytimes.com/2023/04/13/sports/basketball/michael-jordan-retirement-activism.html",

"snippet": "Jordan wasn\u2019t vocal about social justice like today\u2019s N.B.A. stars, but his secret brand of activism is a key reason they have the spotlight now.",

"lead\_paragraph": "On the first road trip of his N.B.A. career, in the fall of 2001, Etan Thomas looked out the window of the Washington Wizards\u2019 team bus and was stunned by the massing crowd around the hotel.",

"print\_section": "A",

"print\_page": "25",

"source": "The New York Times",

"multimedia": [

{

"rank": 0,

"subtype": "xlarge",

"caption": null,

"credit": null,

"type": "image",

"url": "images/2023/04/13/multimedia/13nba-jordan-retirement-01-jczw/13nba-jordan-retirement-01-jczw-articleLarge.jpg",

"height": 400,

"width": 600,

"legacy": {

"xlarge": "images/2023/04/13/multimedia/13nba-jordan-retirement-01-jczw/13nba-jordan-retirement-01-jczw-articleLarge.jpg",

"xlargewidth": 600,

"xlargeheight": 400

},

"subType": "xlarge",

"crop\_name": "articleLarge"

},

}

],

"headline": {

"main": "Michael Jordan Was an Activist After All",

"kicker": "On Pro Basketball",

"content\_kicker": null,

"print\_headline": "It Turns Out That Jordan Was an Activist After All",

"name": null,

"seo": null,

"sub": null

},

"keywords": [

{

"name": "persons",

"value": "Jordan, Michael",

"rank": 1,

"major": "N"

},

{

"name": "subject",

"value": "Basketball",

"rank": 2,

"major": "N"

},

{

"name": "organizations",

"value": "National Basketball Assn",

"rank": 3,

"major": "N"

},

{

"name": "subject",

"value": "Content Type: Personal Profile",

"rank": 4,

"major": "N"

},

{

"name": "subject",

"value": "Philanthropy",

"rank": 5,

"major": "N"

},

{

"name": "subject",

"value": "Black People",

"rank": 6,

"major": "N"

},

{

"name": "subject",

"value": "Blacks",

"rank": 7,

"major": "N"

},

{

"name": "creative\_works",

"value": "Air (Movie)",

"rank": 8,

"major": "N"

},

{

"name": "subject",

"value": "Sneakers",

"rank": 9,

"major": "N"

}

],

"pub\_date": "2023-04-13T09:00:24+0000",

"document\_type": "article",

"news\_desk": "Sports",

"section\_name": "Sports",

"subsection\_name": "Pro Basketball",

"byline": {

"original": "By Harvey Araton",

"person": [

{

"firstname": "Harvey",

"middlename": null,

"lastname": "Araton",

"qualifier": null,

"title": null,

"role": "reported",

"organization": "",

"rank": 1

}

],

"organization": null

},

"type\_of\_material": "News",

"\_id": "nyt://article/3311e67e-a98b-56df-a222-568ed8613faf",

"word\_count": 1711,

"uri": "nyt://article/3311e67e-a98b-56df-a222-568ed8613faf"

},

**Analysis**

Similar fields include: section, headline, date, word count, byline, kicker(NYTimes)/trailText (The Guardian), and web URL. However, there are some fundamental differences about the data returned. For instance, section is a very clear classification schema in the NYTimes print edition and on their website whereas The Guardian has a two-tiered system on its website which is reelected in the combination of the pillar and section. For example, an sectionName might be Business, but the pillarName is News; sectionName Music is pillarName Arts

Also of note is the way that pagination is treated. The NYTimes only allows pagination through pages with 10 results at a time. However, you can loop through the pages by using Python’s time module and math module, keeping in mind that there are two rate limits per API: 500 requests per day and 5 requests per minute. You should sleep 12 seconds between calls to avoid hitting the per minute rate limit. See example query for looping through pages with these modules.

The Guardian allows for between 1-50 results per page to be returned. We added a loop in the code to run the get request for every page of results and aggregate them into a single json file.

**Example Query Comparison**

We requested a list of headlines published December 29, 2022 from both news sources and found the same story to compare:

The Guardian returned the following item in the response to all headlines from items published on December 29, 2022 (see **TheGuardian\_20221229-headlines\_no-API-key.py)** :

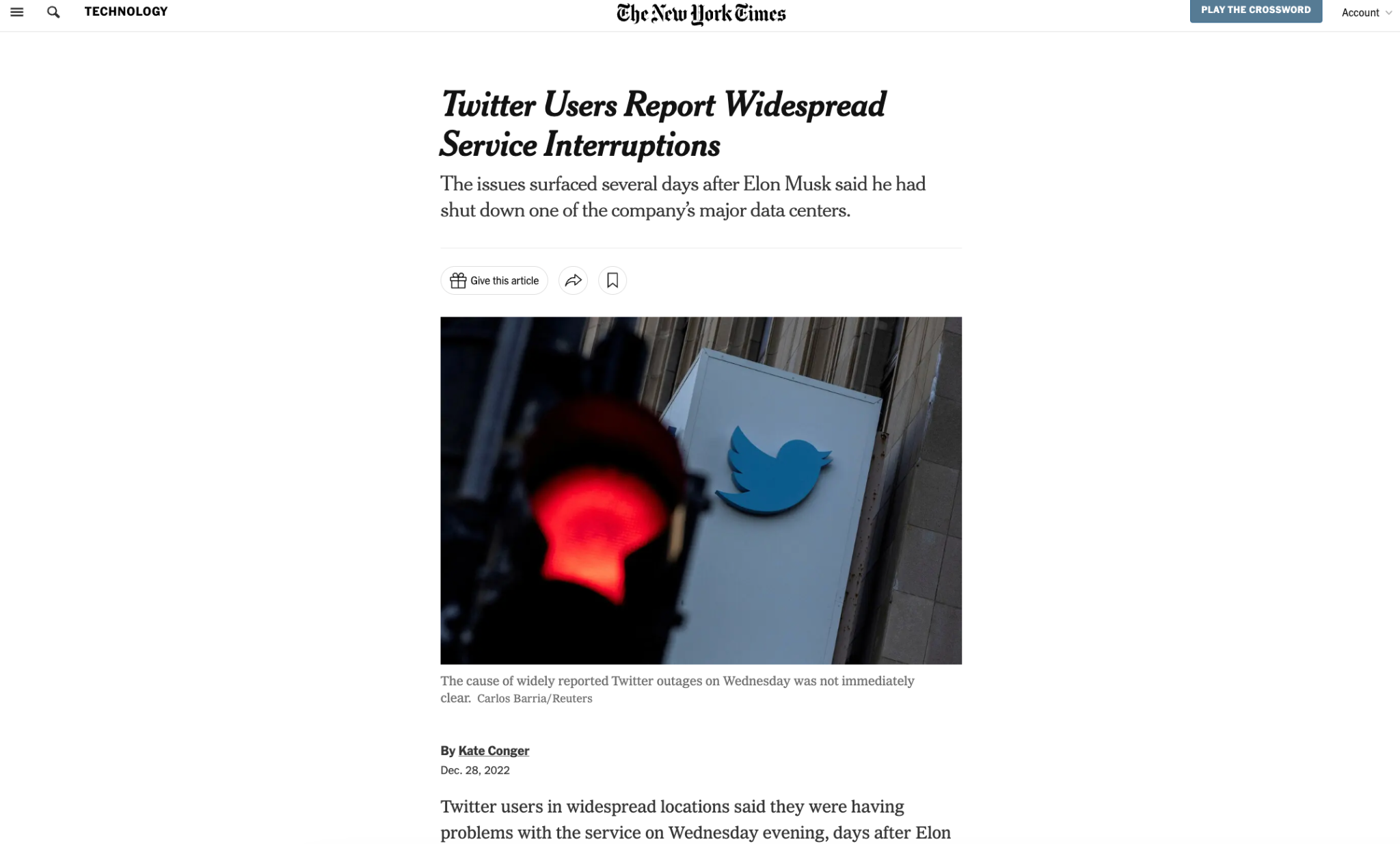
{ "id": "technology/2022/dec/29/twitter-users-report-global-outage-with-many-unable-to-log-into-website-or-app",  
 "type": "article",  
 "sectionId": "technology",  
 "sectionName": "Technology",  
 "webPublicationDate": "2022-12-29T01:34:10Z",  
 "webTitle": "Twitter outage: Elon Musk says \u2018works for me\u2019 as users report problems with website",  
 "webUrl": "https://www.theguardian.com/technology/2022/dec/29/twitter-users-report-global-outage-with-many-unable-to-log-into-website-or-app",  
 "apiUrl": "https://content.guardianapis.com/technology/2022/dec/29/twitter-users-report-global-outage-with-many-unable-to-log-into-website-or-app",  
 "fields": {  
 "headline": "Twitter outage: Elon Musk says \u2018works for me\u2019 as users report problems with website"  
 },  
 "isHosted": false,  
 "pillarId": "pillar/news",  
 "pillarName": "News"  
 },

That item is the following article (found via the webURL field):

The NYTimes returned the following item in the response to all headlines from items published on December 29, 2022 **(see NYTimes\_20221229-headlines.py) :**

{'main': 'Twitter Users Report Widespread Service Interruptions', 'kicker': None, 'content\_kicker': None, 'print\_headline': 'Service Outage at Twitter Appears to Be Far-Reaching; Cause of Disruption Is Not Yet Known', 'name': None, 'seo': None, 'sub': None}

That item is the following article (found via the webURL field): \*\*note this article was published online on December 28th, but was printed in the December 29th issue.



**TAKEAWAYS**

Using Python to parse and query the New York Times and The Guardian’s APIs demonstrates how one can programmatically compare what stories are covered, how the stories are framed and develop over time, and when and how often they’re covered across news sources. As with our example of the Twitter outage (above), the API metadata can reveal both the commonality and differences between the two perspectives and with the code setup, it can easily be adjusted to return different queries and overcome pagination issues to aggregate data in ways that the newspaper websites would struggle to return.

**HOW TO GET STARTED**

Once you’ve requested and received your API keys, you can plug them into the additional sample queries we’ve provided to practice with:

* The above example was part of the data returned by **TheGuardian\_20221229-headlines\_no-API-key.py** and **​​NYTimes\_20221229-headlines.py**These scripts returned all headlines of items published on December 29, 2022.
* **TheGuardian\_women-technology\_no-API-key.py** and **NYTimes\_women\_technology.py**