

# 파이썬 웹 크롤링

---

강사장철원

## □ API 활용해서 뉴스데이터 크롤링하기

# 파이썬 웹 크롤링

Section 1. news api 활용

---

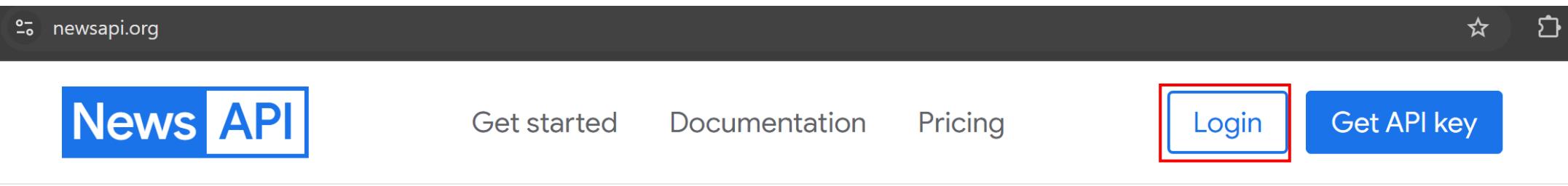
Section 1-1. 크롤링 웹사이트 확인

## Section

### 크롤링 웹사이트 확인

# 크롤링 개요

<https://newsapi.org/>



## Search worldwide news with code

Locate articles and breaking news headlines from news  
sources and blogs across the web with our JSON API

[Get API Key →](#)

## Section

### 크롤링 웹사이트 확인

# 크롤링 개요

## Login

Email address

Password

[Forgotten your password?](#)

No account? [Get one here.](#)

## Register for API key

First name

Email address

Email addresses will be verified, please enter a real one. Disposable addresses have been blocked.

Choose a password



You are...

I am an individual

I am a business, or am working on behalf of a business



로봇이 아닙니다.



reCAPTCHA  
개인정보 보호 - 약관

I agree to the [terms](#).

## Section

### 크롤링 웹사이트 확인

**크롤링 개요** <https://newsapi.org/>

## Registration complete

Your API key is:

`#73F9ad50a7805297669f5aef3376c45`

For help getting started please look at our [getting started guide](#).

We post API status updates and other news on our Twitter feed, so please follow us there if that's important to you:

 Follow @NewsAPIorg

 My account

## Section

### 크롤링 웹사이트 확인

# 크롤링 개요 <https://newsapi.org/>

All articles mentioning Apple from yesterday, sorted by popular publishers first →

GET

<https://newsapi.org/v2/everything?q=apple&from=2024-11-09&to=2024-11-09&sor tBy=popularity&apiKey=d71f9dd5bba74b52976695ae61374c45>

```
{  
  status: "ok",  
  totalResults: 689,  
  - articles: [  
    - {  
      - source: {  
          id: null,  
          name: "[Removed]"  
        },  
        author: null,  
        title: "[Removed]",  
        description: "[Removed]",  
        url: https://removed.com,  
        urlToImage: null,  
        publishedAt: "2024-11-09T15:00:04Z",  
        content: "[Removed]"  
      },  
      - {  
        - source: {  
            id: null,  
            name: "Android Central"  
          }  
        }  
      ]  
    }  
  }  
}
```

## Section

### 크롤링 웹사이트 확인

# 크롤링 개요 <https://newsapi.org/>

```
[1]: from bs4 import BeautifulSoup
from urllib.request import urlopen

[2]: start = "2025-07-20"
apikey = "REDACTED"
keyword = "apple"

[3]: url_popular = "https://newsapi.org/v2/everything?q={}&from={}&sortBy=publishedAt&apiKey={}".format(keyword, start, apikey)

[4]: print(url_popular)

https://newsapi.org/v2/everything?q=apple&from=2025-07-20&sortBy=publishedAt&apiKey=d71f9dd5bba74b52976695ae61374c45

[5]: html = urlopen(url_popular)
bs_obj = BeautifulSoup(html, "html.parser")
print(bs_obj)

{"status": "ok", "totalResults": 6677, "articles": [{"source": {"id": null, "name": "PhoneArena"}, "author": "Tsveta Ermenkova", "title": "Apple stricker", "description": "The tech giant rolls out tougher age labels with iOS 26 beta.", "url": "https://www.phonearena.com/news/stricker_id172615", "urlToImage": "https://m-cdn.phonearena.com/images/article/172615-wide-two_1200/Apples-App-Store-is-getting-stricker": "2025-07-25T06:41:08Z", "content": "To get those new age labels right, Apple is also updating the questionnaire developers have to fill in an app. There are now new required questions covering everything from in-app purchases to privacy settings. [+613 chars]"}, {"source": {"id": null, "name": "Giga.de"}, "author": "Ulfuss", "title": "Kostenlos bei Apple: Dieses unerwartete Geschenk gibt's jetzt für iPhone, iPad und Mac", "description": "Apple überrascht mit einem Geschenk, das du unbedingt sehen musst.", "url": "https://www.giga.de/tech/kostenlos-bei-apple-dieses-unerwartete-geschenk-gibt-s-jetzt-fuer-alle-01K1050JNMPKG3YPVP8ENPHWP9", "urlToImage": "https://static.giga.de/76/e1/da/2662d54a297588b3baa51452d3_cmUgMTkyMCAxMDgwAzcxMjI0OTgt": "2025-07-25T06:41:00Z", "content": "Apple überrascht mit einem Geschenk, das du unbedingt sehen musst.\r\nApple verteilt derzeit ein feines Geschenk an seine Nutzer: individuell gestaltbare Hintergrundbilder, sogenannte Personalized Wallpapers. [+2149 chars]"}, {"source": {"id": null, "name": "Stefan"}, "author": "Stefan", "title": "iPadOS 26 preview: A long-awaited multi-tasking update pays off (so far)", "description": "I'm not going to be able to use it for a while, but I'm excited about the potential of the new features.", "url": "https://www.stefan.it/iPadOS-26-preview-A-long-awaited-multi-tasking-update-pays-off-so-far", "urlToImage": "https://static.stefan.it/2025/07/25/172615-wide-two_1200/iPadOS-26-preview-A-long-awaited-multi-tasking-update-pays-off-so-far.jpg", "publishedAt": "2025-07-25T06:41:08Z", "content": "I'm not going to be able to use it for a while, but I'm excited about the potential of the new features."}]}
```

## Section

## 크롤링 웹사이트 확인

# 크롤링 개요

```
[7]: article_dic = json.loads(bs_obj.text)
      print(article_dic)
```

```
{'status': 'ok', 'totalResults': 6677, 'articles': [{}{'source': {'id': None, 'name': 'PhoneArena'}, 'author': 'Store is getting stricter", 'description': 'The tech giant rolls out tougher age labels with iOS 26 beta.', 'urlToImage': 'https://m-cdn.phonearena.com/images/article/172615-richter.jpg', 'publishedAt': '2025-07-25T06:41:08Z', 'content': 'To get those new age labels right, Apple is have to fill out when submitting an app. There are now new required questions covering everything from in-app purchases, 'name': 'Giga'}, 'author': 'Sven Kaulfuss', 'title': 'Kostenlos bei Apple: Dieses unerwartete Geschenk gibt', 'description': 'Apple überrascht mit einem Geschenk, das du unbedingt sehen musst.', 'url': 'https://www.giga.de/geschenk-gibt-s-jetzt-fuer-iphone-ipad-und-mac--01K1050JNMPKG3YPVP8ENPHWP9', 'urlToImage': 'https://static_cmUgMTkyMCAxMDgwAzcxMjI00TgzYTRk.png', 'publishedAt': '2025-07-25T06:41:00Z', 'content': 'Apple überrascht musst.\r\nApple verteilt derzeit weltweit ein kleines, aber feines Geschenk an seine Nutzer: individuell gehars']}, {'source': {'id': None, 'name': 'Madshrimps.be'}, 'author': 'Stefan', 'title': 'iPadOS 26 preview: first (so far)', 'description': 'I\'m not going to beat around the bush: iPadOS 26 and its new multitasking features are Pretty much ever since Steve Jobs introduced the iPad 15 years ago (!), Apple has tried to straddle two worlds: https://www.madshrimps.be/news/ipados-26-preview-a-long-awaited-multi-tasking-update-pays-off-so-far/', 'urlToImage': 'https://m-cdn.madshrimps.be/images/article/1050JNMPKG3YPVP8ENPHWP9.png', 'publishedAt': '2025-07-25T06:40:35Z', 'content': 'I\'m not going to beat around the bush: iPadOS 26 and its new multitasking features are as much ever since Steve Jobs introduced the iPad 15 years ago (!), Apple ... [+502 chars']}, {'source': {'id': None, 'name': 'TechRadar'}, 'author': 'Sam Bagnall', 'title': 'How to install the iOS 26 public beta', 'description': 'Before we get started:, no, you have lots of major Apple OS updates. In case you missed the news, from now on all of the company\'s various operating systems will receive regular public betas.', 'url': 'https://www.techradar.com/news/how-to-install-the-ios-26-public-beta', 'urlToImage': 'https://m-cdn.techradar.com/images/article/172615-richter.jpg', 'publishedAt': '2025-07-25T06:41:08Z', 'content': 'Before we get started:, no, you have lots of major Apple OS updates. In case you missed the news, from now on all of the company\'s various operating systems will receive regular public betas.'}]}
```

```
[8]: type(article_dic)
```

```
[8]: dict
```

```
[9]: article_dic['articles']
```

```
[9]: [{}{'source': {'id': None, 'name': 'PhoneArena'},  
       'author': 'Tsveta Ermenkova',  
       'title': "Apple's App Store is getting stricter",  
       'description': 'The tech giant rolls out tougher age labels with iOS 26 beta.',  
       'url': 'https://www.phonearena.com/news/apple-app-store-getting-stricter\_id172615',  
       'urlToImage': 'https://m-cdn.phonearena.com/images/article/172615-wide-two\_1200/Apples-App-Store-is-getting-stricter\_id172615.jpg',  
       'publishedAt': '2025-07-25T06:41:08Z',  
       'content': 'To get those new age labels right, Apple is also updating the questionnaire developers have to now new required questions covering everything from in-app co... [+613 chars]'}],
```

## Section

### 크롤링 웹사이트 확인

# 크롤링 개요

<https://newsapi.org/>

```
[10]: article_dic['articles'][0]
```

```
[10]: {'source': {'id': None, 'name': 'PhoneArena'},  
       'author': 'Tsveta Ermenkova',  
       'title': "Apple's App Store is getting stricter",  
       'description': 'The tech giant rolls out tougher age labels with iOS 26 beta.',  
       'url': 'https://www.phonearena.com/news/apple-app-store-getting-stricter_id172615',  
       'urlToImage': 'https://m-cdn.phonearena.com/images/article/172615-wide-two_1200/Apples-App-Store-is-getting-s  
       'publishedAt': '2025-07-25T06:41:08Z',  
       'content': 'To get those new age labels right, Apple is also updating the questionnaire developers have to fi  
       w new required questions covering everything from in-app co... [+613 chars]}'}
```

```
[11]: article_dic['articles'][0]['content']
```

```
[11]: 'To get those new age labels right, Apple is also updating the questionnaire developers have to fill out when  
ed questions covering everything from in-app co... [+613 chars]'
```

```
[12]: len(article_dic['articles'])
```

```
[12]: 95
```

## Section

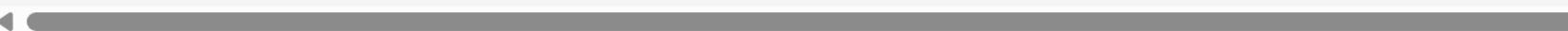
### 크롤링 웹사이트 확인

# 크롤링 개요 <https://newsapi.org/>

```
[13]: n = len(article_dic['articles'])
```

```
[14]: keywords = ['Samsung', 'samsung']
```

```
[15]: a = 'Samsung has been toying with the idea of releasing its own XR (extended reality) headset for around a year. In a recently
```



```
[16]: keywords[0]
```

```
[16]: 'Samsung'
```

```
[17]: keywords[0] in a
```

```
[17]: True
```

```
[18]: keywords = ['Samsung', 'samsung']
count = 0
for i in range(0, n):
    for keyword in keywords:
        if keyword in article_dic['articles'][i]['content']:
            count += 1
            break;
```

```
[19]: count
```

```
[19]: 3
```

감사합니다.

# Q & A