

Trong bài này, chúng ta sẽ thu thập tin tức về dịch Covid-19 và hiển thị thông tin lên giao diện website. Để hiển thị thông tin lên giao diện, chúng ta sẽ tìm hiểu về kiểu dữ liệu từ điển và thực hành tổ chức dữ liệu dưới dạng từ điển, gồm có:

- Từ điển dữ liệu thống kê về dịch Covid
- Danh sách các từ điển về các tin tức, mỗi từ điển lưu một tin tức

### 7.1. Thu thập tin tức về dịch Covid

Các tin tức về dịch Covid sẽ được thu thập tại trang <https://tuoitre.vn/dich-covid-19-e576.htm>. Các thông tin cần được lấy gồm có tiêu đề, mô tả và hình ảnh, chúng ta thực hiện việc lấy mã nguồn và hiển thị như trong Bài 6. Chúng ta thực hiện lại việc lấy thông tin từ trang web mới, các bạn có thể xóa hoặc thêm những lệnh print để hiển thị thông tin tùy ý.

```
...
16 VN_death = list_span[2].text

17 web_driver.get("https://tuoitre.vn/dich-covid-19-e576.htm")
18 html_page = web_driver.page_source

19 def index(request):
20     return HttpResponse(html_page)
```

127.0.0.1:8000

▲ MEDIA THỜI SỰ THẾ GIỚI PHÁP LUẬT KINH DOANH CÔNG NGHỆ XE DU LỊCH NHỊP SỐNG TRẺ VĂN HÓA GIẢI TRÍ THỂ THAO GIÁO DỤC

## Dịch COVID-19

Thông tin về dịch bệnh viêm đường hô hấp cấp COVID-19

**Hôm qua**  
23/01



SỨC KHỎE

**Tin COVID-19 chiều 23-1: Cả nước 14.978 ca nhiễm mới, TP.HCM xuống dưới 200 ca**

TTO - Trong 24 giờ qua, cả nước ghi nhận 14.978 ca nhiễm mới và 4.157 ca khỏi. Hà Nội vẫn có nhiều ca mắc nhất với 2.967. Tại TP.HCM, ca mắc mới giảm xuống chỉ còn 138, tử vong còn 4 ca.

• Tin sáng 23-1: Số ca COVID-19 chuyển nặng phần lớn chưa tiêm vắc xin

**Thứ 5**  
20/01



SỨC KHỎE

**Tin COVID-19 chiều 20-1: Thêm 16.715 ca nhiễm mới, số tử vong nhiều ở miền Tây**

TTO - Cả nước ghi nhận thêm 16.715 ca nhiễm mới trong 24 giờ qua, trong đó Hà Nội 2.884 ca, TP.HCM 245 ca. Có thêm 5.736 bệnh nhân được nhập viện điều trị.



Sau đó, các bạn tìm kiếm các thẻ chứa các thông tin tiêu đề, mô tả và hình ảnh của từng tin tức giống như cách tìm kiếm số liệu về Covid-19 trong Bài 6. Các bạn kiểm tra phần tử sẽ thấy mỗi tin tức nằm trong một thẻ `<li>` với class là "news-item". Các thành phần của tin tức như thời gian, ảnh, tiêu đề... nằm trong các thẻ khác bên trong thẻ `<li>`.

## Dịch COVID-19

Thông tin về dịch bệnh viêm đường hô hấp cấp COVID-19

li.news-item 664 x 149



```

<div class="box-news-latest isstream">
  <ul class="list-news-content">
    <li class="news-item" data-newsid="20220123151324907" data-boxtype="timelineposition" box-type="newstimestreamcate">...</li>
    <li class="news-item" data-newsid="20220120170451788" data-boxtype="timelineposition" box-type="newstimestreamcate">...</li>
    <li class="news-item" data-newsid="202201171746034" data-boxtype="timelineposition" box-type="newstimestreamcate">...</li>
    <li class="news-item" data-newsid="2022011614442188" data-boxtype="timelineposition" box-type="newstimestreamcate">...</li>
  </ul>
</div>

```

Với mỗi tin tức, chúng ta lấy ra các thông tin sau:

- Hình ảnh: Chúng ta có thể lấy được hình ảnh thông qua đường link ảnh là giá trị thuộc tính **src** của thẻ `<img>`.

## Dịch COVID-19

Thông tin về dịch bệnh viêm đường hô hấp cấp COVID-19

img.img212x132 212 x 132

Hôm qua  
24/01



SỨC KHỎE  
Tin COVID-19 ghi nhận ở 1:  
TTO - Bản tin chi  
14.362 ca COVID  
ghi nhận 97 ca m  
• Hậu COVID-19  
tính đực ra se

chieu-24-1-so-ca-nhiem-omicron-ghi-nhan-o-13-tinh-thanh-20220124160010435.htm" data-icon-type="done">0">

```

 == $0

```

Chủ nhật  
23/01



SỨC KHỎE  
Tin COVID-19 nhiễm mới, T

```

<a class="category-name fl mgl10 mg b4 uppercase" href="/suc-khoe.htm" title="Sức khỏe">Sức khỏe</a>

```

- Tiêu đề: Các bạn có thể thấy tiêu đề nằm trong nội dung của thẻ `<a>`, nhưng thực tế trong thẻ `<a>` dùng để chứa đường link, `<li>` sẽ có nhiều thẻ `<a>` chứa các đường link khác nhau, vì vậy ta sẽ không tìm tiêu đề bằng thẻ `<a>`. Ta thấy thẻ `<a>` này nằm trong thẻ tiêu đề `<h3>`, vì vậy chúng ta sẽ tìm kiếm thông qua thẻ `<h3>` này, nội dung của thẻ `<a>` cũng là nội dung của thẻ `<h3>`.

p COVID-19

a 325 x 36



### Tin COVID-19 chiều 24-1: Số ca nhiễm Omicron ghi nhận ở 13 tỉnh thành

TTO - Bản tin chiều 24-1 của Bộ Y tế cho biết cả nước ghi nhận 14.362 ca COVID-19 mới, 163 ca tử vong. Đáng chú ý, TP.HCM chỉ ghi nhận 97 ca mắc trong ngày và 2 ca tử vong, là số mắc và...

• Hậu COVID-19 ảnh hưởng sức khỏe sinh sản, tình dục ra sao?



SỨC KHỎE

### Tin COVID-19 chiều 23-1: Cả nước 14.978 ca nhiễm mới, TP.HCM xuống dưới 200 ca

▼<div class="name-news">

▼<h3 class="title-news">

```
<a title="Tin COVID-19 chiều
24-1: Số ca nhiễm Omicron ghi
nhận ở 13 tỉnh thành" data-
objecttype data-id="20220124
160010435" data-linktype="ne
wsdetail" data-comment="2022
0124160010435" href="/tin-co
vid-19-chieu-24-1-so-ca-nhiem
-omicron-ghi-nhan-o-13-tinh-t
hanh-20220124160010435.htm">
Tin COVID-19 chiều 24-1: Số
ca nhiễm Omicron ghi nhận ở
13 tỉnh thành</a> == $0
</h3>
```

▶<p class="sapo" rel="3">...</p>

- Nội dung tin tức: Nội dung tin tức là nội dung văn bản của thẻ <p>.

SỨC KHỎE

#text 483.83 x 63

### u 24-1: Số ca nhiễm Omicron thành

TTO - Bản tin chiều 24-1 của Bộ Y tế cho biết cả nước ghi nhận 14.362 ca COVID-19 mới, 163 ca tử vong. Đáng chú ý, TP.HCM chỉ ghi nhận 97 ca mắc trong ngày và 2 ca tử vong, là số mắc và...

• Hậu COVID-19 ảnh hưởng sức khỏe sinh sản, tình dục ra sao?

SỨC KHỎE

### Tin COVID-19 chiều 23-1: Cả nước 14.978 ca nhiễm mới, TP.HCM xuống dưới 200 ca

▼<div class="name-news">

▶<h3 class="title-news">...</h3>

▼<p class="sapo" rel="3">

"TTO - Bản tin chiều 24-1 của Bộ Y tế cho biết cả nước ghi nhận 14.362 ca COVID-19 mới, 163 ca tử vong. Đáng chú ý, TP.HCM chỉ ghi nhận 97 ca mắc trong ngày và 2 ca tử vong, là số mắc và..."

</p>

▶<div class="news-extra-one no-thumb">...</div>

Chúng ta sẽ thực hiện phân tích cú pháp và lấy về danh sách các thẻ <li> với class là "news-item" chứa các tin tức (ví dụ: **articles\_html**). Các bạn có thể hiển thị ra màn hình kích thước và nội dung danh sách để kiểm tra trường hợp: Ngoài hiển thị tin tức ra thẻ còn sử dụng để hiển thị nội dung khác hay không. Các bạn lưu ý toàn bộ nội dung các thẻ tin tức sẽ rất dài, các bạn có thể lựa chọn kiểm tra bằng cách hiển thị phần tử đầu, phần tử cuối, phần tử ngẫu nhiên bất kỳ... trong danh sách tin tức.

```
17 web_driver.get("https://tuoitre.vn/dich-covid-19-e576.htm")
18 html_page = web_driver.page_source

19 parser = BeautifulSoup(html_page, 'html.parser')
20 articles_html = parser.find_all('li', class_='news-item')
21 print('-----')
22 print(len(articles_html))
23 print(articles_html)
24 print('-----')
```





```

Terminal: Local X Local (2) X Local (3) X + v
855e0a9&rc=s (2)
-----
so luong: 20
[<li box-type="newstimelinestreamcate" class="news-item" data-boxtype="timelineposition" data-newsid
="20220124160010435">
<div class="published-date fl">
<span class="first-label clearfix">Hôm qua</span>
<span class="second-label" title="">24/01</span>
</div>
<a class="img212x132 pos-rlt" data-icon-type-done="0" data-objecttype="" href="/tin-covid-19-chieu-2
4-1-so-ca-nhiem-omicron-ghi-nhan-o-13-tinh-thanh-20220124160010435.htm" title="Tin COVID-19 chiều 24
-1: Số ca nhiễm Omicron ghi nhận ở 13 tỉnh thành">
` đầu tiên của phần tử đầu tiên
- `articles_html[0].a.img['src']` trả về giá trị thuộc tính **src** của thẻ `<img>` đầu tiên trong thẻ `<a>` của phần tử đầu tiên.
- `title = articles_html[0].h3.text` trả về nội dung của thẻ `<h3>` đầu tiên của phần tử đầu tiên
- `title = articles_html[0].p.text` trả về nội dung của thẻ `<p>` đầu tiên của phần tử đầu tiên

Như vậy ta đã tách được các thông tin cần tìm từ tin tức đầu tiên.

```

...
20 articles_html = parser.find_all('li', class_='news-item')

21 img_src = articles_html[0].a.img['src']
22 title = articles_html[0].h3.text
23 content = articles_html[0].p.text
24 print('-----')
25 print(img_src)
26 print(title)
27 print(content)

```







SỨC KHỎE

### Tin COVID-19 chiều 24-1: Số ca nhiễm Omicron ghi nhận ở 13 tỉnh thành

TTO - Bản tin chiều 24-1 của Bộ Y tế cho biết cả nước ghi nhận 14.362 ca COVID-19 mới, 163 ca tử vong. Đáng chú ý, TP.HCM chỉ ghi nhận 97 ca mắc trong ngày và 2 ca tử vong, là số mắc và...

- Hậu COVID-19 ảnh hưởng sức khỏe sinh sản, tình dục ra sao?

```
Terminal: Local x Local (2) x Local (3) x + v
-----
https://cdn.tuoiitre.vn/zoom/188_117/2022/1/24/z309213088021130fb65d8969bab1393063dc5a8ca3816-1643013
8840861386753065-crop-16430145673881119031116.jpg

Tin COVID-19 chiều 24-1: Số ca nhiễm Omicron ghi nhận ở 13 tỉnh thành

TTO - Bản tin chiều 24-1 của Bộ Y tế cho biết cả nước ghi nhận 14.362 ca COVID-19 mới, 163 ca tử von
g. Đáng chú ý, TP.HCM chỉ ghi nhận 97 ca mắc trong ngày
và 2 ca tử vong, là số mắc và...

System check identified no issues (0 silenced).
```

## 7.2. Hiện thị thông tin lên website

Hiện tại mỗi phần tử của danh sách `articles_html` là một tin tức với định dạng là mã HTML, chúng ta đã tách được các thông tin phần tử đầu tiên thành các dữ liệu đường link hình ảnh, tiêu đề và nội dung. Để lấy được thông tin của toàn bộ danh sách tin tức, chúng ta có thể có các cách tổ chức dữ liệu như sau:

- Cách 1: Sử dụng 3 danh sách, một danh sách chứa toàn bộ link hình ảnh của các tin tức, mỗi phần tử là link hình ảnh của một tin tức; một danh sách chứa các tiêu đề; một danh sách chứa các nội dung.
- Cách 2: Sử dụng một danh sách chứa các tin tức, mỗi phần tử là một danh sách chứa 3 thông tin link hình ảnh, tiêu đề, nội dung của một tin tức.
- Cách 3: Sử dụng một danh sách chứa các tin tức, mỗi phần tử là một từ điển chứa 3 thông tin link hình ảnh, tiêu đề, nội dung của một tin tức.

Chúng ta sẽ sử dụng cách thứ 3. Kiểu dữ liệu từ điển (Dictionary) là một cấu trúc lưu trữ dữ liệu dưới dạng cặp khóa và giá trị (key – value). Trong kiểu dữ liệu từ điển các khóa sẽ có giá trị duy nhất (không có hai khóa nào giống nhau) và thứ tự các cặp khóa - giá trị là không quan trọng. Từ điển được khai báo với cặp dấu ngoặc nhọn bao ngoài, bên trong các phần tử theo cặp khóa - giá trị được phân tách bởi dấu phẩy, giữa khóa và giá trị của từng phần tử được phân cách bởi dấu hai chấm.

```
dict = {khóa_1 : giá trị_1, khóa_2 : giá trị_2, ...}
```

Ví dụ:

```
student = {'name': 'Thanh Bình', 'age': "12"}
```



Để truy xuất giá trị của các phần tử, từ điển sẽ sử dụng cặp dấu ngoặc vuông và truy xuất thông qua khóa (key) thay vì số thứ tự như danh sách. Ví dụ `student['name']` sẽ có giá trị là 'Thanh Binh'.

Mỗi tin tức sẽ được lưu là một từ điển với 3 phần tử là 3 thông tin cần lưu. Các bạn tạo một từ điển (ví dụ: **article\_dict**) và có thể đặt tên các khóa cố định là **img\_src**, **title** và **content**, các bạn lưu ý việc đặt tên khóa này là bắt buộc theo quy ước với dữ liệu trong tập tin **web\_covid.html**, giá trị lấy từ các biến đã được thiết lập từ trước.

```
...
20 articles_html = parser.find_all('li', class_='news-item')

21 img_src = articles_html[0].a.img['src']
22 title = articles_html[0].h3.text
23 content = articles_html[0].p.text

24 article_dict = {'img_src': img_src, 'title': title,
                  'content': content}
25 print('-----')
26 print(article_dict)
```

```
Terminal: Local x Local (2) x Local (3) x + v
=s (2)
-----
{'img_src': 'https://cdn.tuoiitre.vn/zoom/188_117/2022/1/24/z309213088021130fb65d8969bab1393063dc5a8ca3816-16430138840861386753065-crop-16430145673881119031116.jpg', 'title': '\nTin COVID-19 chiều 24-1: Số ca nhiễm Omicron ghi nhận ở 13 tỉnh thành\n', 'content': 'TTT - Bản tin chiều 24-1 của Bộ Y tế cho biết c'
à nước ghi nhận 14.362 ca COVID-19 mới, 163 ca tử vong. Đáng chú ý, TP.HCM chỉ ghi nhận 97 ca mắ
c trong ngày và 2 ca tử vong, là số mắc và...'}
System check identified no issues (0 silenced).
You have 18 unapplied migration(s). Your project may not work properly until you apply the migra
```

Các lệnh từ dòng 21 – 26 trên thực hiện tạo dữ liệu từ điển cho tin tức đầu tiên, tuy nhiên thực tế chúng ta cần một danh sách chứa các từ điển là toàn bộ các tin tức, mỗi từ điển biểu diễn một tin tức. Để có được danh sách các từ điển này, chúng ta tạo một danh sách rỗng (ví dụ: **article\_list\_dict**) như dòng 21 và sử dụng vòng lặp duyệt danh sách **articles\_html**, với mỗi phần tử trong danh sách chúng ta tạo thành một từ điển và thêm từ điển đó vào **article\_list** bằng lệnh **append()** như ví dụ trong dòng 27.

```
...
20 articles_html = parser.find_all('li', class_='news-item')
21 article_list_dict = []
22 for article in articles_html:
23     img_src = article.a.img['src']
24     title = article.h3.text
```



```

25 content = article.p.text
26 article_dict = {'img_src': img_src, 'title': title,
                  'content': content}
27 article_list_dict.append(article_dict)
28 print('-----')
29 print(article_list_dict)

```

Sau khi vòng lặp kết thúc, **article\_list** là danh sách các từ điển về các tin tức mà chúng ta cần. Chúng ta bổ sung thêm từ điển dữ liệu thống kê về dịch Covid-19 (ví dụ: **covid\_stats\_dict**) với các khóa bắt buộc là "VN\_case", "VN\_recovery" và "VN\_death".

```

... ..
21 article_list_dict = []
22 for article in articles_html:
... ..

28 covid_stat_dict = {"VN_case": VN_case, "VN_recovery":
                    VN_recovery, "VN_death": VN_death}

```

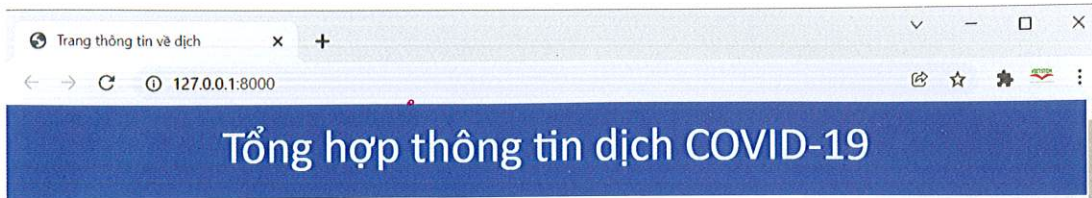
Cuối cùng, chúng ta trả về kết quả chứa nội dung là tập tin **web\_covid.html** kèm với một từ điển với hai khóa bắt buộc "covid\_stat" và "article\_list" với các giá trị đã được thiết lập bên trên để hiển thị trên trình duyệt.

```

28 covid_stat_dict = {"VN_case": VN_case, "VN_recovery":
                    VN_recovery, "VN_death": VN_death}
29 def index(request):
30     return render(request, 'web_covid.html',
                    {'covid_stat': covid_stat_dict,
                     'article_list': article_list_dict})

```

Các bạn chạy dự án và đã thấy dữ liệu được điền vào trang chủ. Để hiển thị được dữ liệu, tệp tin mã nguồn **web\_covid.html** đã sử dụng Jinja, đây là phần kiến thức nâng cao và sẽ không được giới thiệu chi tiết ở đây, các bạn có thể tìm hiểu thêm thông qua từ khóa "Jinja template".



Nguồn số liệu: Báo VNExpress. Link: <https://vnexpress.net/covid-19/covid-19-viet-nam>

#### Việt Nam

**TỔNG SỐ CA NHIỄM**  
2.149.011

**KHỎI**  
1.904.069

**TỬ VONG**  
37.013

#### Thế giới

**TỔNG SỐ CA NHIỄM**

**KHỎI**

**TỬ VONG**

Nguồn tin tức: Báo Tuổi trẻ. Link: <https://tuoitre.vn/dich-covid-19-e576.htm>



Nguồn tin tức: Báo Tuổi trẻ. Link: <https://tuoitre.vn/dich-covid-19-e576.htm>

## Tin tức



### Tin COVID-19 chiều 24-1: Số ca nhiễm Omicron ghi nhận ở 13 tỉnh thành

TTO - Bản tin chiều 24-1 của Bộ Y tế cho biết cả nước ghi nhận 14.362 ca COVID-19 mới, 163 ca tử vong. Đáng chú ý, TP.HCM chỉ ghi nhận 97 ca mắc trong ngày và 2 ca tử vong, là số mắc và...



### Tin COVID-19 chiều 23-1: Cả nước 14.978 ca nhiễm mới, TP.HCM xuống dưới 200 ca

TTO - Trong 24 giờ qua, cả nước ghi nhận 14.978 ca nhiễm mới và 4.157 ca khỏi. Hà Nội vẫn có nhiều ca mắc nhất với 2.967. Tại TP.HCM, ca mắc mới giảm xuống chỉ còn 138, tử vong còn 4 ca.



### Tin COVID-19 chiều 20-1: Thêm 16.715 ca nhiễm mới, số tử vong nhiều ở miền Tây





Các bạn có thể thấy phần thông tin thống kê về dịch Covid-19 trên thế giới vẫn đang để trống, việc lập trình lấy dữ liệu và hiển thị các thông tin này là bài tập thực hành cho các bạn.



### Tóm tắt lý thuyết và bài tập thực hành

Trong bài này, chúng ta sẽ thu thập tin tức về dịch Covid-19 và tìm hiểu về kiểu dữ liệu từ điển, qua đó hiển thị thông tin lên giao diện website.

**Bài tập 1.** Trang web nguồn chúng ta lấy tin tức dịch covid là trang web nào?

**Bài tập 2.** Mỗi tin tức chúng ta cần lấy những thông tin gì?

**Bài tập 3.** Lập trình hiển thị số liệu số ca nhiễm, số ca khỏi và số ca tử vong trên thế giới với một nguồn dữ liệu bất kỳ các bạn tìm được, ví dụ số liệu về số ca nhiễm và số ca tử vong có thể tham khảo tại địa chỉ: <https://vnexpress.net/covid-19/covid-19-the-gioi/>