



WEB CRAWLER

Scrapy Spider

MARCOS FOLGUERA RIVERA

Código Python

```
import scrapy
```

```
import csv
```

```
class RtveSpider(scrapy.Spider):
```

```
    name = 'rtve_spider'
```

```
    start_urls = ['https://www.rtve.es/noticias/']
```

```
    def parse(self, response):
```

```
        # Seleccionar las noticias usando XPath
```

```
        news = response.xpath('//a[span[@class="maintitle"]]')
```

```
        # Escribir las noticias en un archivo CSV
```

```
        with open('rtve_news.csv', 'w', newline='') as file:
```

```
            writer = csv.writer(file)
```

```
            for new in news:
```

```
                title = new.xpath('span[@class="maintitle"]/text()').get()
```

```
                url = new.xpath('@href').get()
```

```
                writer.writerow([title, url])
```

```
import scrapy
import csv

class RtveSpider(scrapy.Spider):
    name = 'rtve_spider'
    start_urls = ['https://www.rtve.es/noticias/']

    def parse(self, response):
        # Seleccionar las noticias usando XPath
        news = response.xpath('//a[span[@class="maintitle"]]')

        # Escribir las noticias en un archivo CSV
        with open('rtve_news.csv', 'w', newline='') as file:
            writer = csv.writer(file)
            for new in news:
                title = new.xpath('span[@class="maintitle"]/text()').get()
                url = new.xpath('@href').get()
                writer.writerow([title, url])
```

Nuestro programa recopila los títulos y su url de las noticias de rtve.es/noticias y los almacena en un fichero .csv

Adjunto captura donde se pueden apreciar como obtiene las noticias

```
GNU nano 2.3.1 Fichero: rtve_news.csv
Los barcos camuflados de Rusia: vigilancia oculta para planificar sabotajes en el Mar del Norte,https://www.rtve.es/noticias/20230419/rusia-barcos-arrastre-sabotaje-ma$
Joe Biden invita a Pedro Sánchez a la Casa Blanca el próximo 12 de mayo en el inicio de la campaña electoral,https://www.rtve.es/noticias/20230419/joe-biden-invita-ped$
Juan Carlos I llega a Sanxenxo para participar en las regatas en su segunda visita a España,https://www.rtve.es/noticias/20230419/rey-emerito-regresa-espana-casi-ano-d$
El Banco de España rebaja a unos 200.000 los beneficiarios de las medidas de alivio hipotecario,https://www.rtve.es/noticias/20230419/banco-espana-hipotecas-economia-i$
Sánchez anuncia la financiación de 43.000 viviendas de alquiler social a precios asequibles,https://www.rtve.es/noticias/20230419/sanchez-financiacion-viviendas/243956$
"Planas advierte de la ""situación inédita"" de sequía: ""Hay que adaptarse al cambio climático que tenemos ante nosotros""",https://www.rtve.es/noticias/20230419/plan$
El Real Betis anuncia la retirada de Joaquín a final de temporada,https://www.rtve.es/deportes/20230419/joaquin-betis-retiro-futbol/2439697.shtml
"Cuenta atrás para la primera vacuna de ARN mensajero contra el cáncer: ""Salvará millones de vidas""",https://www.rtve.es/noticias/20230419/vacuna-arn-mensajero-modern$
EE.UU. anuncia un nuevo paquete de ayuda militar para Ucrania de 325 millones de dólares,https://www.rtve.es/noticias/20230420/minuto-minuto-guerra-ucrania/2439701.sht$
Podcast 'Diario de Ucrania': las otras guerras de Putin,https://www.rtve.es/noticias/20230419/mark-galeotti-ucrania-guerras-putin/2439448.shtml
Mapa: Rusia bombardea 112 localidades ucranianas y avanza sobre Bajmut,https://www.rtve.es/noticias/20230419/mapa-guerra-ucrania/2438755.shtml
"Albares, sobre los españoles atrapados en Sudán: ""Los combates en torno al aeropuerto nos impide sacarlos""",https://www.rtve.es/noticias/20230419/albares-espanoles-$
España estudia más ayudas para paliar los efectos de la sequía en el campo y pedirá flexibilidad a Bruselas,https://www.rtve.es/noticias/20230419/espana-estudia-ayudas$
La segunda visita del rey emérito a Sanxenxo genera división de opiniones en las distintas fuerzas políticas,https://www.rtve.es/noticias/20230419/reacciones-visita-re$
"Belarra ve ""electoralismo"" en los anuncios: ""Para que se cumplan Podemos tiene que gobernar con fuerza""",https://www.rtve.es/noticias/20230419/belarra-vivienda-el$
"El PP tacha de ""bluf"" los anuncios de Sánchez sobre vivienda: ""No tiene ninguna credibilidad""",https://www.rtve.es/noticias/20230419/pp-bluf-anuncios-sanchez-vivi$
"La oposición arremete contra la ""sumisión"" de Sánchez frente a Marruecos y este les reta a explicar una ""política alternativa""",https://www.rtve.es/noticias/20230$
Los funcionarios de Justicia se manifiestan en Madrid para exigir mejoras laborales y salariales,https://www.rtve.es/noticias/20230419/huelga-manifestacion-funcionario$
"El Gobierno analizará el ""fichaje"" en Arabia Saudí de investigadores españoles para subir en los rankings",https://www.rtve.es/noticias/20230419/universidades-anali$
Un policía investigado por narcotráfico muere en un tiroteo con otros agentes cuando iban a detenerlo,https://www.rtve.es/noticias/20230419/muere-policia-nacional-tiro$
Ocho sociedades científicas piden retirar las mascarillas en hospitales y centros sanitarios,https://www.rtve.es/noticias/20230419/mascarilla-centros-sanitarios-socied$
Nueve de cada diez jóvenes están expuestos en redes sociales al consumo de tabaco o vapeadores,https://www.rtve.es/noticias/20230419/jovenes-expuestos-redes-tabaco-vap$
"Las lluvias regresan este fin de semana, previo a días de calor de pleno verano",https://www.rtve.es/noticias/20230419/lluvias-regresan-fin-semana-previo-dias-calor-p$
Sanidad retira del mercado los potenciadores sexuales 'Mero Macho' y 'Mero Macho Premium',https://www.rtve.es/noticias/20230419/sanidad-retira-del-mercado-potenciadores$
"Los pilotos de Air Europa convocan una huelga para los días 1,2, 4 y 5 de mayo",https://www.rtve.es/noticias/20230419/pilotos-air-europa-convocan-huelga-para-dias-12-$
"La inflación frenó al 6,9% en la eurozona, pero la subyacente marcó un récord del 5,7% en marzo",https://www.rtve.es/noticias/20230419/inflacion-freno-69-eurozona-per$
"La marca de moda de lujo Gucci, investigada por la Comisión Europea en el marco de una operación antimonopolio",https://www.rtve.es/noticias/20230419/gucci-investigad$
"Facua demanda a diez fabricantes de coches por inflar precios, en representación de 766 compradores",https://www.rtve.es/noticias/20230419/facua-demanda-diez-fabrican$
"La luz cae más de un 4% hasta los 108,3 euros/MWh este jueves",https://www.rtve.es/noticias/20230419/luz-cae-mas-4-ciento-hasta-1083-euros-mwh-este-jueves/2439631.sht$
El líder del mayor sindicato francés y rostro de la lucha por las pensiones dimitirá en junio,https://www.rtve.es/noticias/20230419/lider-del-mayor-sindicato-frances-r$
"El Ejército y las milicias de la FAR vuelven a vulnerar el alto el fuego por segundo día en Sudán,https://www.rtve.es/noticias/20230419/continuan-enfrentamientos-sudan$
"El bombero que capturó el levantamiento del gueto de Varsovia: ""Observaba detrás de los muros y sacaba las fotos en secreto""",https://www.rtve.es/noticias/20230419/$
Dos adolescentes detenidos y acusados de asesinato por el tiroteo en una fiesta de cumpleaños en Alabama,https://www.rtve.es/noticias/20230419/dos-adolescentes-detenid$
Un juez de EE.UU. ordena detener al expresidente peruano Alejandro Toledo para extraditarlo a Perú,https://www.rtve.es/noticias/20230420/juez-eeuu-ordena-detener-expre$
¿Es la ayuda humanitaria altruista o egoísta?,https://www.rtve.es/noticias/20230419/trabajadores-humanitarios-deseo-ayudar-projimo-confundir-egoismo/2432055.shtml
La 'tremdina' del Ramadán: mal humor e irritabilidad como consecuencia del ayuno,https://www.rtve.es/noticias/20230419/marruecos-tremdina-ramadan-consecuencias-ayuno/2$
"La educación de las mujeres iraníes está en riesgo por los envenenamientos, según Amnistía Internacional",https://www.rtve.es/noticias/20230419/educacion-mujeres-iran$
"Díaz-Canel, reelegido como presidente de Cuba para un segundo mandato",https://www.rtve.es/noticias/20230419/secretario-del-partido-comunista-cuba-diaz-canel-reelegid$
Nicolás Maduro cumple 10 años al frente de Venezuela,https://www.rtve.es/play/videos/telediario-2/nicolas-maduro-diez-anos-frente-venezuela/6867866/
```

← → ↺ rtve.es/noticias/ Campus Correo unex YouTube WhatsApp Twitch de jajas HITBOXING Pelis Series Peliculas miedo Dios PelisSeries Pendiente Uni LINER: ChatGP

DIRECTO: rtve Noticias Televisión Radio Deportes Infantil RTVEPlay PlayRadio ElTiempo Playz



Joe Biden invita a Pedro Sánchez a la Casa Blanca el próximo 12 de mayo en el inicio de la campaña electoral



Juan Carlos I llega a Sanxenxo para participar en las regatas en su



Los barcos camuflados de Rusia: vigilancia oculta para planificar sabotajes en el Mar del Norte

Una investigación de cuatro cadenas públicas nórdicas revela actividades de vigilancia encubiertas de Rusia.

Webgrafía:

<https://docs.scrapy.org/en/latest/intro/tutorial.html>

<https://chat.openai.com/>

<https://elmundodelosdatos.com/extraccion-datos-sitios-web-productos-zara/>