



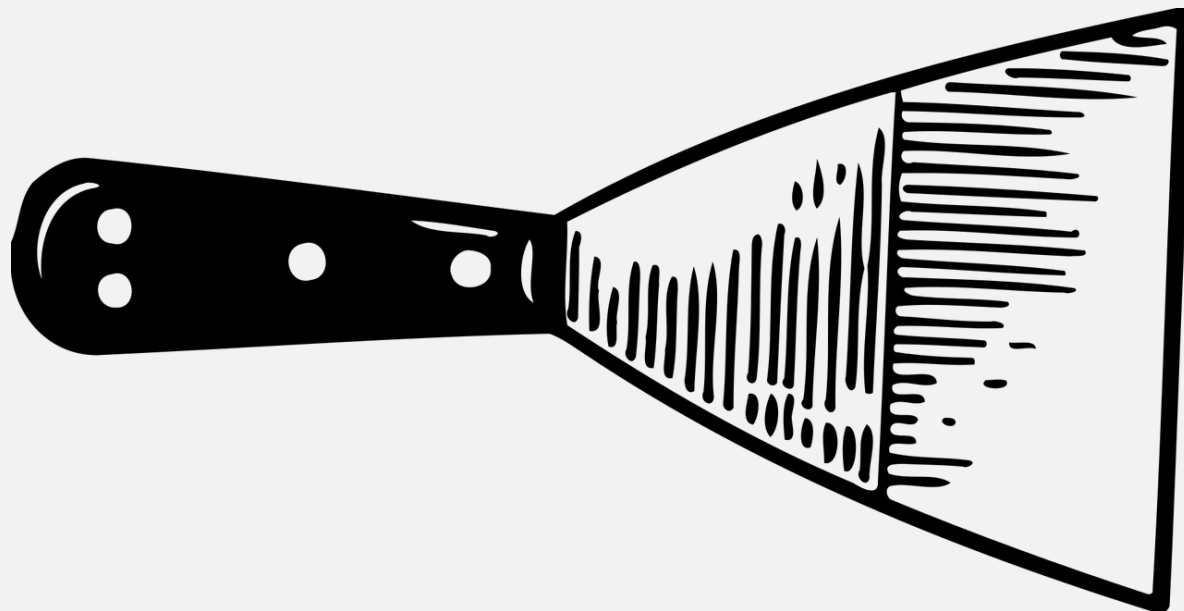
WEB SCRAPING SIMPLE Y AVANZADO

Jose Vicente Yago Martinez



VicenteYago/UMUR-webScraping





¿QUE ES EL WEB SCRAPING?



PNY Nvidia Quadro RTX 4000 8 GB GDDR6

969,90€

Recibelo el **martes 23 de marzo**

4.3 ★ 3 Opiniones

☐ Comparar



PNY Nvidia Quadro P4000 8GB GDDR5

884,90€

Recibelo entre el **lunes 29 de marzo** y el **jueves 1 de abril**

4.8 ★ 11 Opiniones

☐ Comparar



Gigabyte GeForce RTX 2080 SUPER GAMING OC Waterforce WB 8GB GDDR6

619,35€

Recibelo el **miércoles 24 de marzo**

5 ★ 2 Opiniones

☐ Comparar



PNY Nvidia Quadro RTX 4000 8 GB GDDR6
969,90€
Recibelo el martes 23 de marzo
4.3 ★ 3 Opiniones
☐ Comparar




PNY Nvidia Quadro P4000 8GB GDDR5
884,90€
Recibelo entre el lunes 29 de marzo y el jueves 1 de abril
4.8 ★ 11 Opiniones
☐ Comparar



Gigabyte GeForce RTX 2080 SUPER GAMING OC Waterforce WB 8GB GDDR6
619,35€
Recibelo el miércoles 24 de marzo
5 ★ 2 Opiniones
☐ Comparar

Nombre	Inicio	Fin	Dias	Kc
-	1	7	7	0.50
-	8	14	7	0.60
-	15	21	7	0.65
-	22	28	7	0.68
-	29	35	7	0.70
-	36	42	7	0.73
-	43	49	7	0.75
-	50	56	7	0.78
-	57	63	7	0.80

Examples of DNA viruses					
Virus family	Examples (common names)	Virion naked/enveloped	Capsid symmetry	Nucleic acid type	Group
1. <i>Adenoviridae</i>	Canine hepatitis virus, Some types of the common cold	Naked	Icosahedral	ds	I
2. <i>Papovaviridae</i>	JC virus, HPV	Naked	Icosahedral	ds circular	I
3. <i>Parvoviridae</i>	Human parvovirus B19, canine parvovirus	Naked	Icosahedral	ss	II
4. <i>Herpesviridae</i>	Herpes simplex virus, varicella-zoster virus, cytomegalovirus, Epstein–Barr virus	Enveloped	Icosahedral	ds	I
5. <i>Poxviridae</i>	Smallpox virus, cowpox, myxoma virus, monkeypox, vaccinia virus	Complex coats	Complex	ds	I
6. <i>Anelloviridae</i>	Torque teno virus	Naked	Icosahedral	ss circular	II
7. <i>Pleolipoviridae</i>	HHPV1, HRPV1	Enveloped		ss/ds linear/circular	I/II



PNY Nvidia Quadro RTX 4000 8 GB GDDR6
969,90€
Recibelo el **martes 23 de marzo**
4.3 ★ 3 Opiniones
☐ Comparar



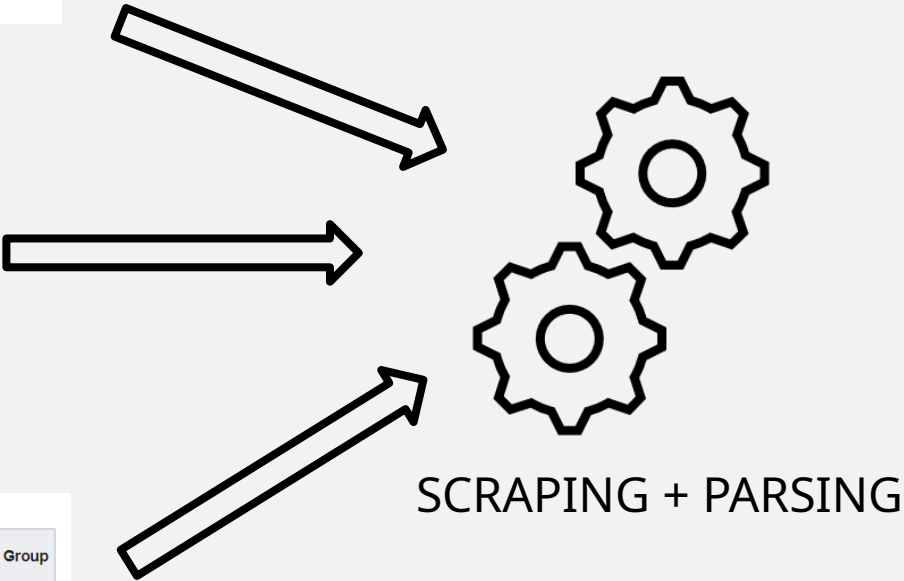
PNY Nvidia Quadro P4000 8GB GDDR5
884,90€
Recibelo entre el **lunes 29 de marzo** y el **jueves 1 de abril**
4.8 ★ 11 Opiniones
☐ Comparar



Gigabyte GeForce RTX 2080 SUPER GAMING OC Waterforce WB 8GB GDDR6
619,35€
Recibelo el **miércoles 24 de marzo**
5 ★ 2 Opiniones
☐ Comparar

Nombre	Inicio	Fin	Dias	Kc
-	1	7	7	0.50
-	8	14	7	0.60
-	15	21	7	0.65
-	22	28	7	0.68
-	29	35	7	0.70
-	36	42	7	0.73
-	43	49	7	0.75
-	50	56	7	0.78
-	57	63	7	0.80

Examples of DNA viruses					
Virus family	Examples (common names)	Virion naked/enveloped	Capsid symmetry	Nucleic acid type	Group
1. <i>Adenoviridae</i>	Canine hepatitis virus, Some types of the common cold	Naked	Icosahedral	ds	I
2. <i>Papovaviridae</i>	JC virus, HPV	Naked	Icosahedral	ds circular	I
3. <i>Parvoviridae</i>	Human parvovirus B19, canine parvovirus	Naked	Icosahedral	ss	II
4. <i>Herpesviridae</i>	Herpes simplex virus, varicella-zoster virus, cytomegalovirus, Epstein-Barr virus	Enveloped	Icosahedral	ds	I
5. <i>Poxviridae</i>	Smallpox virus, cowpox, myxoma virus, monkeypox, vaccinia virus	Complex coats	Complex	ds	I
6. <i>Anelloviridae</i>	Torque teno virus	Naked	Icosahedral	ss circular	II
7. <i>Pleolipoviridae</i>	HHPV1, HRPV1	Enveloped		ss/ds linear/circular	I/II





PNY Nvidia Quadro RTX 4000 8 GB GDDR6
969,90€
Recibelo el martes 23 de marzo
4.3 ★ 3 Opiniones
☐ Comparar



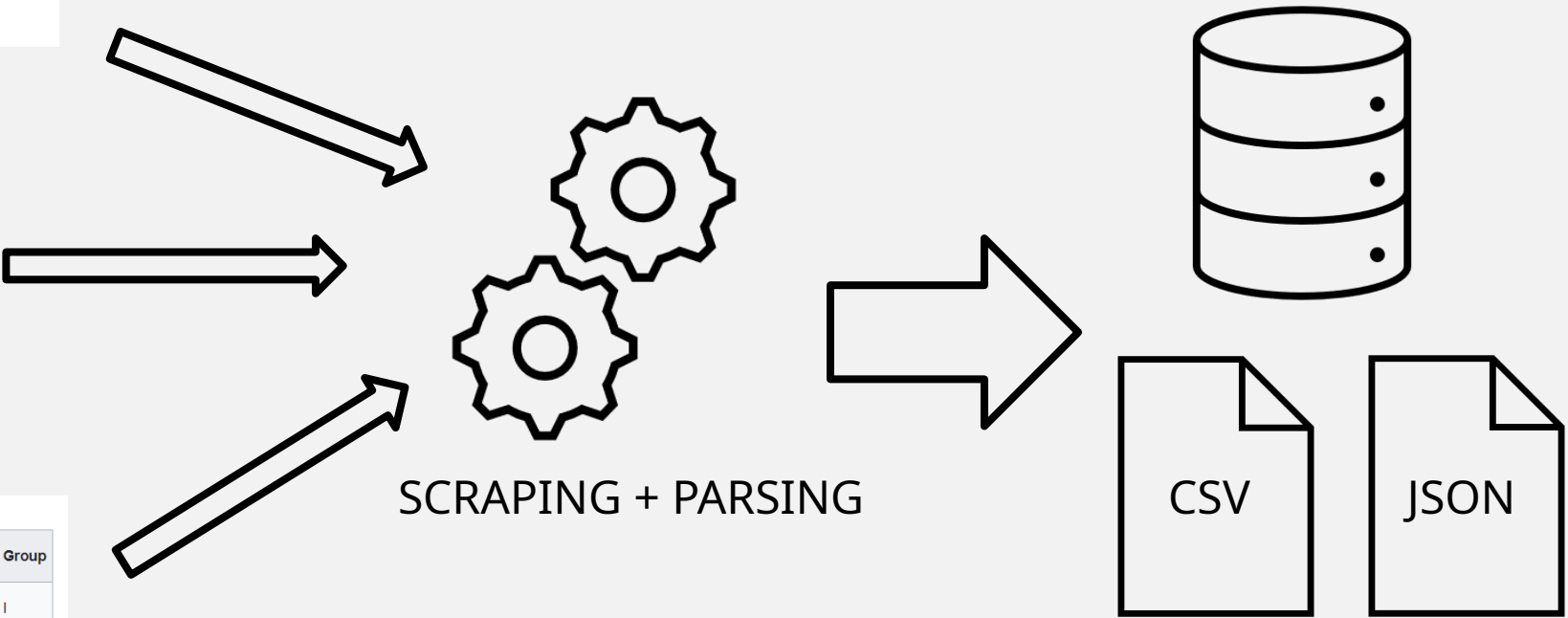
PNY Nvidia Quadro P4000 8GB GDDR5
884,90€
Recibelo entre el lunes 29 de marzo y el jueves 1 de abril
4.8 ★ 11 Opiniones
☐ Comparar



Gigabyte GeForce RTX 2080 SUPER GAMING OC Waterforce WB 8GB GDDR6
619,35€
Recibelo el miércoles 24 de marzo
5 ★ 2 Opiniones
☐ Comparar

Nombre	Inicio	Fin	Dias	Kc
-	1	7	7	0.50
-	8	14	7	0.60
-	15	21	7	0.65
-	22	28	7	0.68
-	29	35	7	0.70
-	36	42	7	0.73
-	43	49	7	0.75
-	50	56	7	0.78
-	57	63	7	0.80

Examples of DNA viruses					
Virus family	Examples (common names)	Virion naked/enveloped	Capsid symmetry	Nucleic acid type	Group
1. <i>Adenoviridae</i>	Canine hepatitis virus, Some types of the common cold	Naked	Icosahedral	ds	I
2. <i>Papovaviridae</i>	JC virus, HPV	Naked	Icosahedral	ds circular	I
3. <i>Parvoviridae</i>	Human parvovirus B19, canine parvovirus	Naked	Icosahedral	ss	II
4. <i>Herpesviridae</i>	Herpes simplex virus, varicella-zoster virus, cytomegalovirus, Epstein-Barr virus	Enveloped	Icosahedral	ds	I
5. <i>Poxviridae</i>	Smallpox virus, cowpox, myxoma virus, monkeypox, vaccinia virus	Complex coats	Complex	ds	I
6. <i>Anelloviridae</i>	Torque teno virus	Naked	Icosahedral	ss circular	II
7. <i>Pleolipoviridae</i>	HHPV1, HRPV1	Enveloped		ss/ds linear/circular	I/II





COMPONENTES DE UNA WEB


```
<!DOCTYPE html>
<html>
<head>
<style>
table, th, td {
  border: 1px solid black;
}
</style>
</head>

  <title>H</title>
  <h3>HEAD: Minichara UMUR: Scraping Simple y Avanzado</h3>
  <p>¡Espero que la disfruteis!</p>

<body>

  <h3>BODY: Tabla sencilla y links</h2>

  <table style="width:100%">
    <tr>
      <th>Firstname</th>
      <th>Lastname</th>
      <th>Age</th>|
    </tr>
    <tr>
      <td>Jill</td>
      <td>Smith</td>
      <td>50</td>
    </tr>
    <tr>
      <td>Eve</td>
      <td>Jackson</td>
      <td>94</td>
    </tr>
    <tr>
      <td>John</td>
      <td>Doe</td>
      <td>80</td>
    </tr>
  </table>

  <p>No te olvides de visitar <a href="https://rvest.tidyverse.org/">Rvest</a>.</p>

</body>
</html>
```

HEAD: Minichara UMUR: Scraping Simple y Avanzado

¡Espero que la disfruteis!

BODY: Tabla sencilla y links

Firstname	Lastname	Age
Jill	Smith	50
Eve	Jackson	94
John	Doe	80

No te olvides de visitar [Rvest](https://rvest.tidyverse.org/).

```
<!DOCTYPE html>
<html>
<head>
<style>
table, th, td {
  border: 1px solid black;
}
</style>
</head>

  <title>H</title>
  <h3>HEAD: Minichara UMUR: Scraping Simple y Avanzado</h3>
  <p>¡Espero que la disfruteis!</p>

<body>

  <h3>BODY: Tabla sencilla y links</h2>

  <table style="width:100%">
    <tr>
      <th>Firstname</th>
      <th>Lastname</th>
      <th>Age</th>|
    </tr>
    <tr>
      <td>Jill</td>
      <td>Smith</td>
      <td>50</td>
    </tr>
    <tr>
      <td>Eve</td>
      <td>Jackson</td>
      <td>94</td>
    </tr>
    <tr>
      <td>John</td>
      <td>Doe</td>
      <td>80</td>
    </tr>
  </table>

  <p>No te olvides de visitar <a href="https://rvest.tidyverse.org/">Rvest</a>.</p>

</body>
</html>
```

HEAD: Minichara UMUR: Scraping Simple y Avanzado

¡Espero que la disfruteis!

BODY: Tabla sencilla y links

Firstname	Lastname	Age
Jill	Smith	50
Eve	Jackson	94
John	Doe	80

No te olvides de visitar [Rvest](https://rvest.tidyverse.org/).

```
<!DOCTYPE html>
<html>
<head>
<style>
table, th, td {
  border: 1px solid black;
}
</style>
</head>

<title>H</title>
<h3>HEAD: Minichara UMUR: Scraping Simple y Avanzado</h3>
<p>¡Espero que la disfruteis!</p>

<body>

<h3>BODY: Tabla sencilla y links</h2>

<table style="width:100%">
<tr>
<th>Firstname</th>
<th>Lastname</th>
<th>Age</th>|
</tr>
<tr>
<td>Jill</td>
<td>Smith</td>
<td>50</td>
</tr>
<tr>
<td>Eve</td>
<td>Jackson</td>
<td>94</td>
</tr>
<tr>
<td>John</td>
<td>Doe</td>
<td>80</td>
</tr>
</table>

<p>No te olvides de visitar <a href="https://rvest.tidyverse.org/">Rvest</a>.</p>

</body>
</html>
```

HEAD: Minichara UMUR: Scraping Simple y Avanzado

¡Espero que la disfruteis!

BODY: Tabla sencilla y links

Firstname	Lastname	Age
Jill	Smith	50
Eve	Jackson	94
John	Doe	80

No te olvides de visitar [Rvest](https://rvest.tidyverse.org/).

```
<!DOCTYPE html>
<html>
<head>
<style>
table, th, td {
  border: 1px solid black;
}
</style>
</head>

<title>H</title>
<h3>HEAD: Minichara UMUR: Scraping Simple y Avanzado</h3>
<p>¡Espero que la disfruteis!</p>

<body>

<h3>BODY: Tabla sencilla y links</h2>

<table style="width:100%">
  <tr>
    <th>Firstname</th>
    <th>Lastname</th>
    <th>Age</th>
  </tr>
  <tr>
    <td>Jill</td>
    <td>Smith</td>
    <td>50</td>
  </tr>
  <tr>
    <td>Eve</td>
    <td>Jackson</td>
    <td>94</td>
  </tr>
  <tr>
    <td>John</td>
    <td>Doe</td>
    <td>80</td>
  </tr>
</table>

<p>No te olvides de visitar <a href="https://rvest.tidyverse.org/">Rvest</a>.</p>

</body>
</html>
```

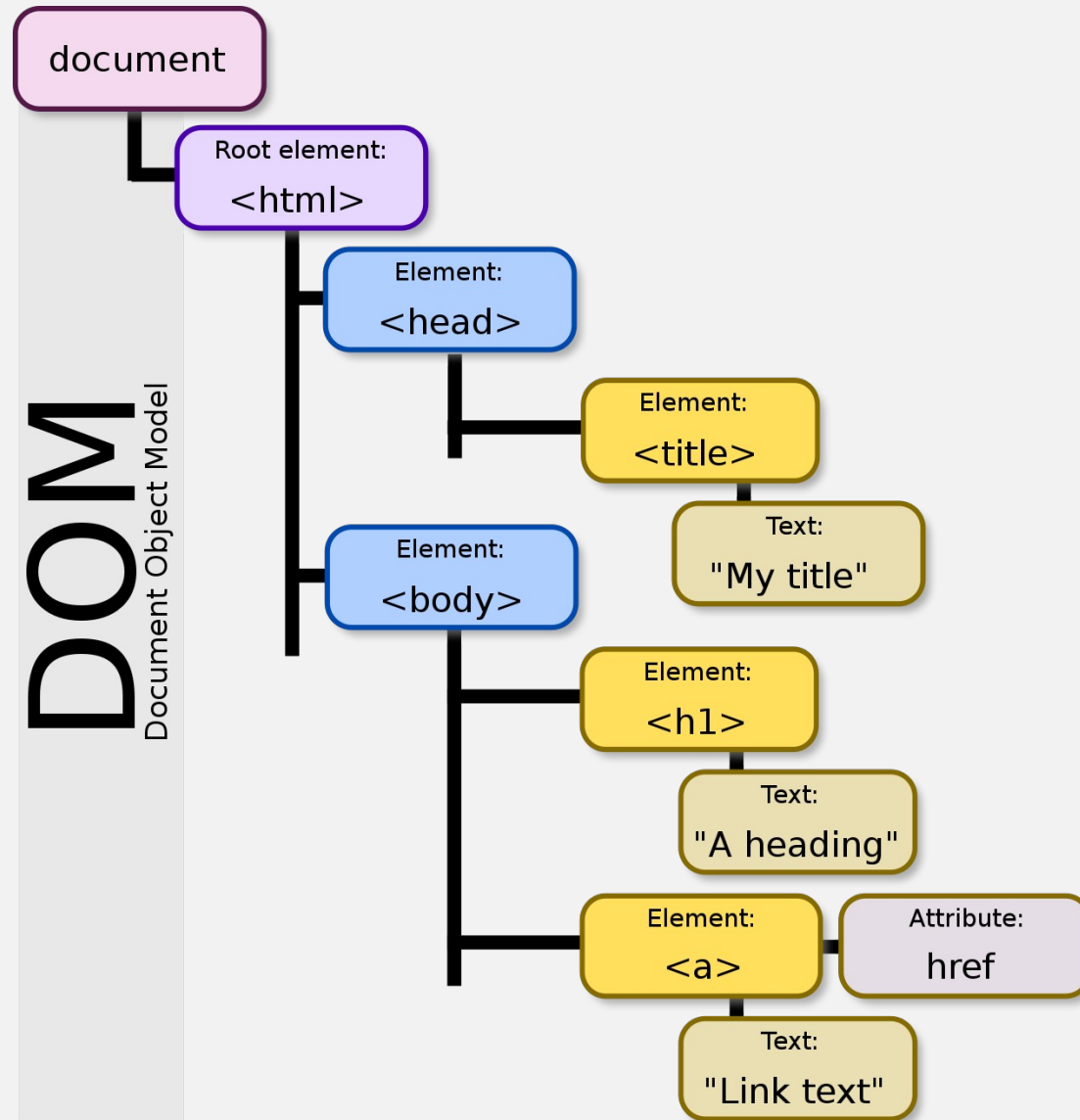
HEAD: Minichara UMUR: Scraping Simple y Avanzado

¡Espero que la disfruteis!

BODY: Tabla sencilla y links

Firstname	Lastname	Age
Jill	Smith	50
Eve	Jackson	94
John	Doe	80

No te olvides de visitar [Rvest](https://rvest.tidyverse.org/).



PAQUETES RELEVANTES EN R



- WEBS ESTÁTICAS
- 80% DE LOS CASOS
- FACIL INSTALACIÓN Y USO



- WEBS COMPLEJAS CON CONTENIDO DINÁMICO
- 20% DE LOS CASOS
- LA INSTALACIÓN SE PUEDE COMPLICAR



SCRAPING SIMPLE



Sony PlayStation 5 Digital + Sony DualSense Mando Inalámbrico para PS5

469,90€ SIN IVA 388,35€

★★★★★ 7 Opiniones | [Review](#)

Vendido y enviado por **PcComponentes** ¿Qué es esto?

Marca: Sony - P/N: SPSSDS | Cod. Artículo: 363741

Envío: Desde 5,25€ **GRATIS** con PcComponentes Premium

Cantidad:

- 1 +

Disponibilidad: Sin fecha exacta de entrada

Financiación :

Cetelem De 3 a 40 meses [+ info](#)

Aplazame De 2 a 30 meses (inmediata) [+ info](#)

Avísame

Sony PlayStation 5 Digital + Sony DualSense Mando Inalámbrico para PS5 Ver más opiniones 6
opiniones

“Que la gente sea un poco más paciente no ay stock en toda España un poco más de paciencia porfavor”

Meek
05/03/2021
★★★★★

Ventajas:
Desventajas : Por ahora ninguna



Sony PlayStation 5 Digital + Sony DualSense Mando Inalámbrico para PS5

469,90€ SIN IVA 388,35€ 

★★★★☆ 7 Opiniones | [Review](#)

Vendido y enviado por **PcComponentes** ¿Qué es esto?

Marca: Sony - P/N: SPSSDS | Cod. Artículo: 363741


Envío: Desde 5,25€ **GRATIS** con PcComponentes Premium

Cantidad:

Disponibilidad: Sin fecha exacta de entrada

Financiación:

 Cetelem De 3 a 40 meses [+ info](#)


 Aplazame De 2 a 30 meses (inmediata) [+ info](#)

  [Avísame](#)

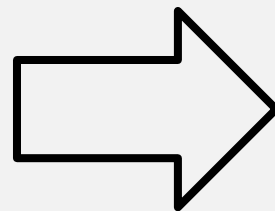
Sony PlayStation 5 Digital + Sony DualSense Mando Inalámbrico para PS5 opiniones

[Ver más opiniones 6](#)

 "Que la gente sea un poco más paciente no ay stock en toda España un poco más de paciencia porfavor"

Meek **Ventajas:**
05/03/2021 **Desventajas:** Por ahora ninguna

★★★★☆



```
Elements Console Sources Network Performance Memory Application Security Lighthouse
<script></script>
</div>
<div class="col-xs-12 col-sm-12 col-md-6"></div>
<script type="application/json" id="microdata-product-script">
{
  "@context": "http://www.schema.org",
  "@type": "product",
  "image": "/thumb.pccomponentes.com/w-530-530/articles/36/363741/1446-sony-playstation-5-digital-sony-dualsense-mando-inalambrico-para-ps5.jpg",
  "name": "Sony PlayStation 5 Digital + Sony DualSense Mando Inalámbrico para PS5",
  "productId": "SPSSDS",
  "sku": "363741",
  "url": "https://www.pccomponentes.com/sony-playstation-5-digital-sony-dualsense-mando-inalambrico-para-ps5",
  "description": "",
  "brand": "Sony",
  "category": "Videoconsolas PS5",
  "itemCondition": "https://schema.org/NewCondition",
  "model": "SPSSDS",
  "mpn": "SPSSDS",
  "aggregateRating": {
    "@type": "AggregateRating",
    "itemReviewed": "https://www.pccomponentes.com/sony-playstation-5-digital-sony-dualsense-mando-inalambrico-para-ps5",
    "ratingValue": "3",
    "reviewCount": "7",
    "ratingCount": "7"
  },
  "offers": {
    "@type": [
      "Offer",
      "PriceSpecification"
    ],
    "price": "469.9",
    "priceCurrency": "EUR",
    "priceSpecification": {
      "@type": "DeliveryChargeSpecification",
      "price": "5.25"
    }
  },
  "availability": "http://schema.org/OutOfStock",
  "url": "https://www.pccomponentes.com/sony-playstation-5-digital-sony-dualsense-mando-inalambrico-para-ps5",
  "itemOffered": "https://www.pccomponentes.com/sony-playstation-5-digital-sony-dualsense-mando-inalambrico-para-ps5",
  "offeredBy": "PcComponentes",
  "priceValidUntil": "2021-03-24"
},
  "review": {
    "@type": "Review",
    "author": {
      "@type": "Person",
      "name": "Meek"
    },
    "datePublished": "2021-03-05",
    "reviewBody": "Que la gente sea un poco más paciente no ay stock en toda España un poco más de paciencia porfavor",
    "reviewRating": {
      "@type": "Rating",
      "ratingValue": "5",
      "bestRating": "5",
      "worstRating": "0"
    }
  },
  "@type": "Review",
  "author": {
    "@type": "Person",
    "name": "anonimo"
  },
  "datePublished": "2021-03-02",
  "reviewBody": "A ver... Eno de como unos pocos es mentira. hav que ser realista v si no hav stock en otras tiendas de videojuegos..."
}
... 363741 div.pccom-super-sliders.slider-ficha-producto div.pccom-super-slider-big.white-card-movil.m-b-0.owl-carousel.owl-theme div.owl-wrapper-outer div.owl-wrapper div.owl-item div.item ...
ld+json 3 of 4 Cancel
```



Sony PlayStation 5 Digital + Sony DualSense Mando Inalámbrico para PS5

469,90€ SIN IVA 388,35€ 

★★★★☆ 7 Opiniones | [Review](#)

Vendido y enviado por **PcComponentes** [¿Qué es esto?](#)


Marca: Sony - P/N: SPSSDS | Cod. Artículo: 363741


Envío: Desde 5,25€ **GRATIS** con PcComponentes Premium

Cantidad:

Disponibilidad: Sin fecha exacta de entrada

Financiación:

 **Cetelem** De 3 a 40 meses [+ info](#)

 **Aplazame** De 2 a 30 meses (inmediata) [+ info](#)



Avísame



Sony PlayStation 5 Digital + Sony DualSense Mando Inalámbrico para PS5 opiniones

[Ver más opiniones 6](#)



"Que la gente sea un poco más paciente no ay stock en toda España un poco más de paciencia porfavor"

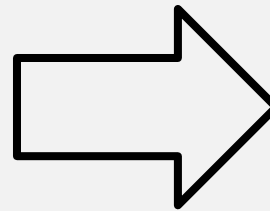
Meek

Ventajas:

05/03/2021

Desventajas: Por ahora ninguna

★★★★☆



Elements Console Sources Network Performance Memory Application Security Lighthouse

</script></script>

</div>

<div class="col-xs-12 col-sm-12 col-md-6"></div>

<script type="application/ld+json" id="microdata-product-script">

{

"@context": "http://www.schema.org",

"@type": "product",

"image": "/thumb.pccomponentes.com/w-530-530/articles/36/363741/1446-sony-playstation-5-digital-sony-dualsense-mando-inalambrico-para-ps5.jpg",

"name": "Sony PlayStation 5 Digital + Sony DualSense Mando Inalámbrico para PS5",

"productID": "SPSSDS",

"sku": "363741",

"url": "https://www.pccomponentes.com/sony-playstation-5-digital-sony-dualsense-mando-inalambrico-para-ps5",

"description": "",

"brand": "Sony",

"category": "Videoconsolas PS5",

"itemCondition": "https://schema.org/NewCondition",

"model": "SPSSDS",

"mpn": "SPSSDS",

"aggregateRating": {

"@type": "AggregateRating",

"itemReviewed": "https://www.pccomponentes.com/sony-playstation-5-digital-sony-dualsense-mando-inalambrico-para-ps5",

"ratingValue": "3",

"reviewCount": "7",

"ratingCount": "7"

},

"offers": {

"@type": ["Offer", "PriceSpecification"],

"price": "469.9",

"priceCurrency": "EUR",

"priceSpecification": {

"@type": "DeliveryChargeSpecification",

"price": "5.25"

},

"availability": "http://schema.org/OutOfStock",

"url": "https://www.pccomponentes.com/sony-playstation-5-digital-sony-dualsense-mando-inalambrico-para-ps5",

"itemOffered": "https://www.pccomponentes.com/sony-playstation-5-digital-sony-dualsense-mando-inalambrico-para-ps5",

"offeredBy": "PcComponentes",

"priceValidUntil": "2021-03-24"

},

"review": {

{

"@type": "Review",

"author": {

"@type": "Person",

"name": "Meek"

},

"datePublished": "2021-03-05",

"reviewBody": "Que la gente sea un poco más paciente no ay stock en toda España un poco más de paciencia porfavor",

"reviewRating": {

"@type": "Rating",

"ratingValue": "5",

"bestRating": "5",

"worstRating": "0"

},

"review": {

"@type": "Review",

"author": {

"@type": "Person",

"name": "anonimo"

},

"datePublished": "2021-03-02",

"reviewBody": "A ver... Es de como unos pocos es mentira... hay que ser realista y si no hay stock en otras tiendas... de videojuegos...

},

},

},

}

}</script>

... 363741 div.pccom-super-sliders.slider-ficha-producto div.pccom-super-slider-big.white-card-movil.m-b-0.owl-carousel.owl-theme div.owl-wrapper-outer div.owl-wrapper div.owl-item div.item ...

ld+json 3 of 4 Cancel

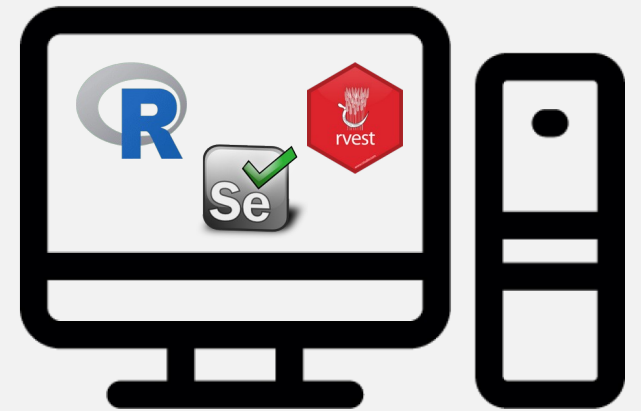
```
'''{r}
scrapeJson('https://www.pccomponentes.com/sony-playstation-5-digital-sony-dualsense-mando-inalambrico-para-ps5',
            '//*[@id="microdata-product-script"]/text()') -> ps5
```

```
sprintf("PRECIO : %s %s", ps5$offers$price, ps5$offers$priceCurrency)
sprintf("DISPONIBLE? : %s", ps5$offers$availability)
sprintf("COMENTARIO 1: %s", ps5$review$reviewBody[1])
'''
```

```
[1] "PRECIO : 469.9 EUR"
[1] "DISPONIBLE? : http://schema.org/OutOfStock"
[1] "COMENTARIO 1: Que la gente sea un poco más paciente no ay stock en toda España un poco más de paciencia porfavor"
```



SCRAPING AVANZADO

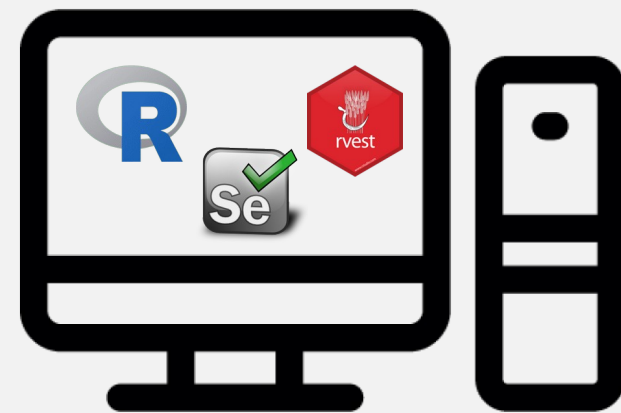
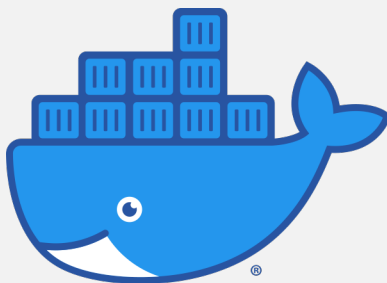


Google

Buscar con Google

Voy a tener suerte

Ofrecido por Google en: [English](#) [català](#) [galego](#) [euskara](#)



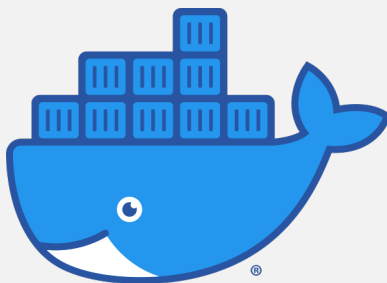
Google



Buscar con Google

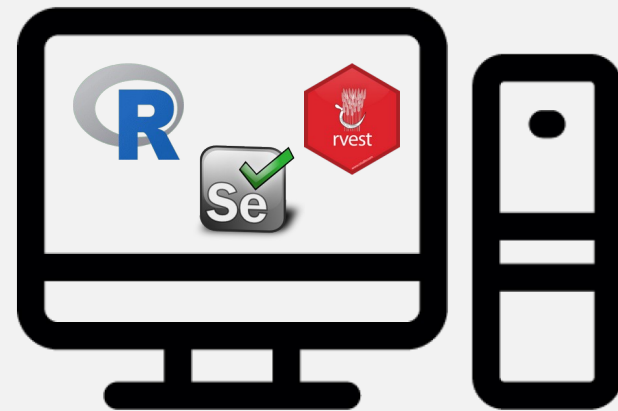
Voy a tener suerte

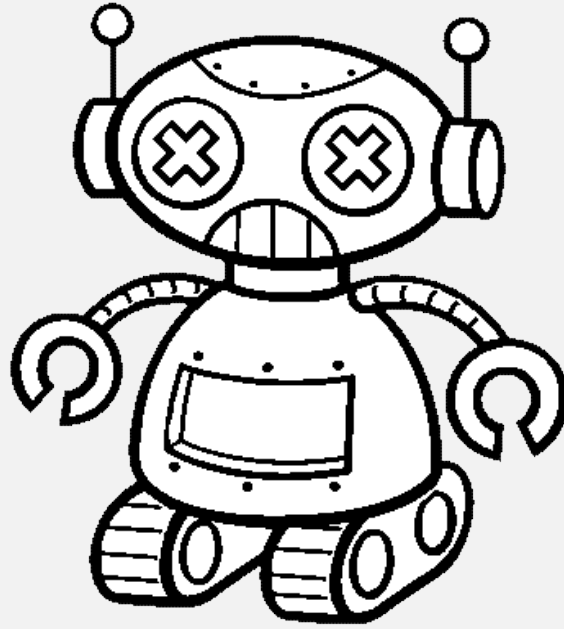
Ofrecido por Google en: [English](#) [català](#) [galego](#) [euskara](#)



ORDENES

RESULTADOS





RESUMEN

- La parte visible de una página web es solo la punta del iceberg.
- Desde R podemos acceder a las *tripas* de cualquier web.
- Rvest es la mejor herramienta para recuperar elementos simples
- Rselenium nos permite manejar el navegador de forma remota
- Con estas 2 herramientas nos podemos crear:
 - Datasets
 - Diversos bots: Alertas, compras automáticas



¡GRACIAS!

