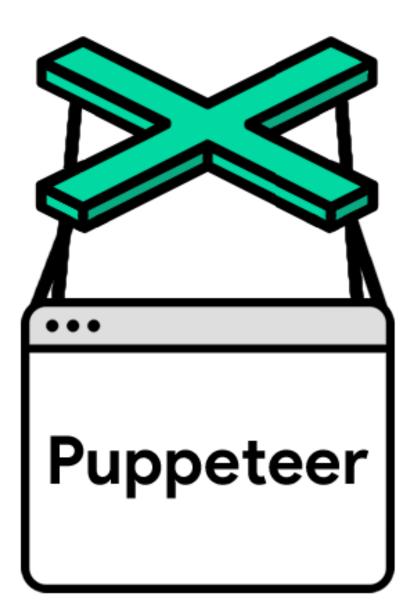
WEB SCRAPPING



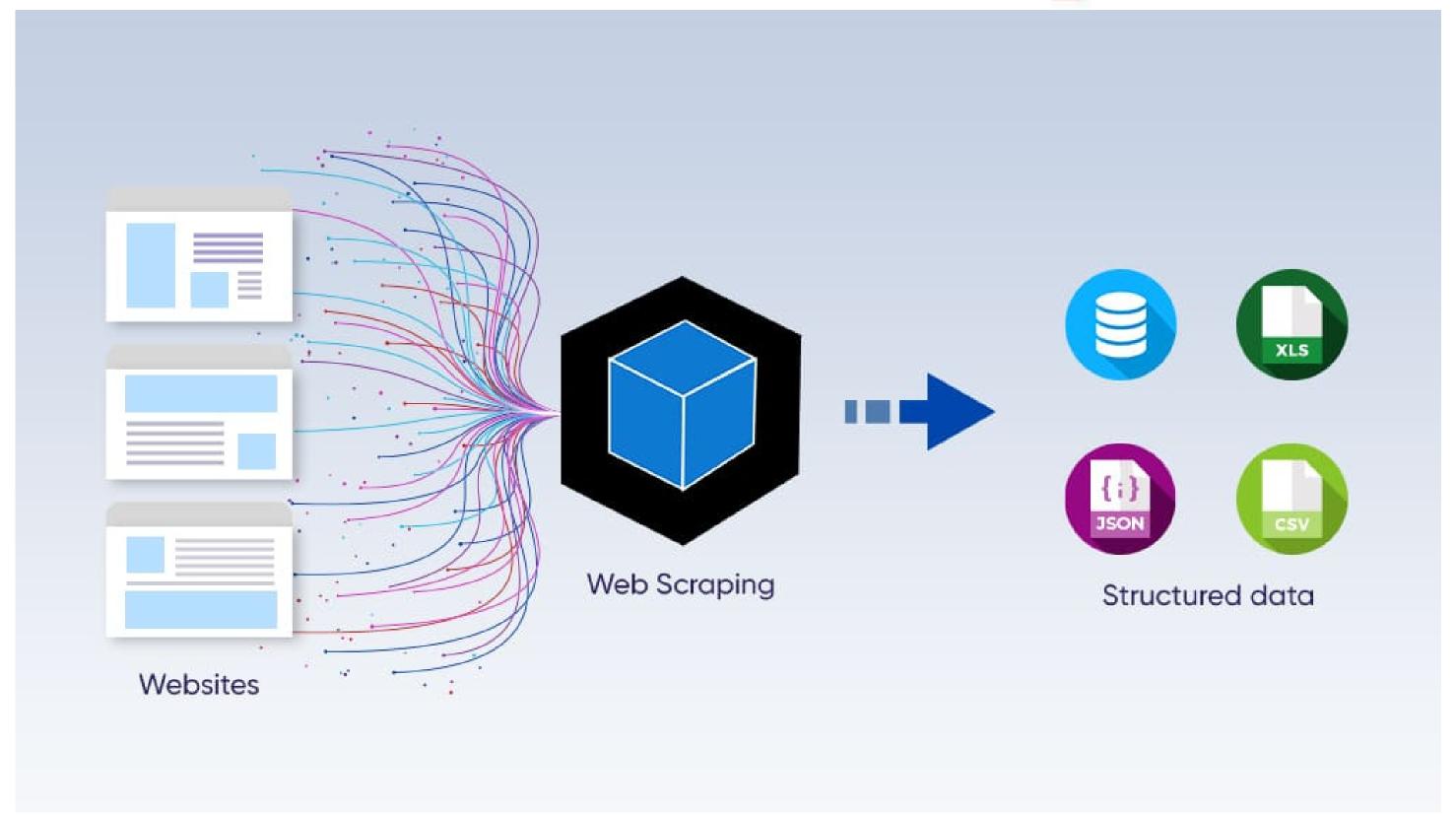


¿QUÉ ES WEB SCRAPPING?









SELECTORES



img.thumbnail 106.34 × 155

Background □#FFFFFF

Margin 0px 43.8594px

Padding 4px

ACCESSIBILITY

Name Thirst

Role image

Keyboard-focusable ○

```
Keyboard-focusable

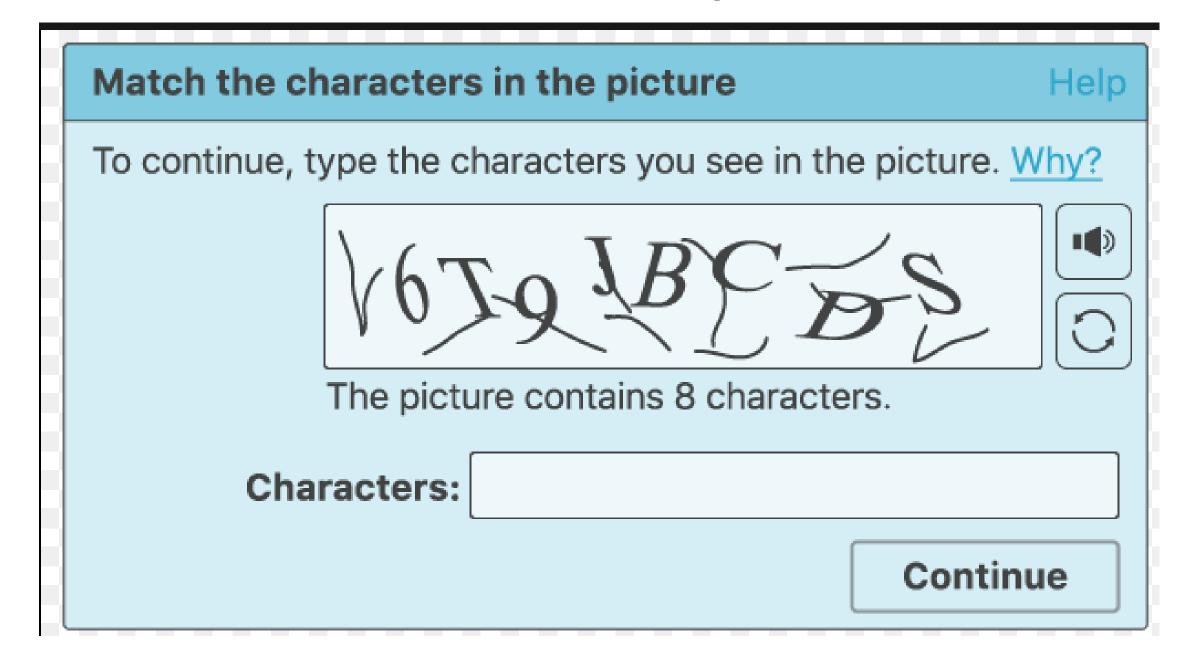
THIRST
```

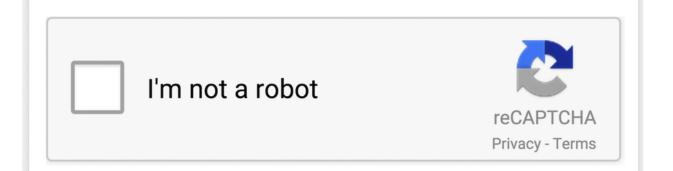
```
▼
  ::before
 \(\text{li class="col-xs-6 col-sm-4 col-md-3 col-lg-3">\(\cdots\)
 \(\text{li class="col-xs-6 col-sm-4 col-md-3 col-lg-3">\(\cdot\)
 ▶  ···· 
 ▼
  ▼ ⟨article class="product_pod"⟩
   ▼ <div class="image container">
    ▼ <a href="../../thirst 946/index.html">
       <img src="../../../media/cache/c4/0a/c40a64f....jpg" alt="Thirst"</pre>
       class="thumbnail">
      </a>
    </div>
   ▶ class="star-rating Five"> ••• 
   ▶ <div class="product_price"> • </div>
   </article>
  ▶  ···· 
 ▶  ···· 
 \(\text{li class="col-xs-6 col-sm-4 col-md-3 col-lg-3">\(\text{...}\)
 ▶  ···· 
 ▶/li class="college_6 college_4 college_3 college_3"\.......//li\
                               inner ul.breadcrumb
```

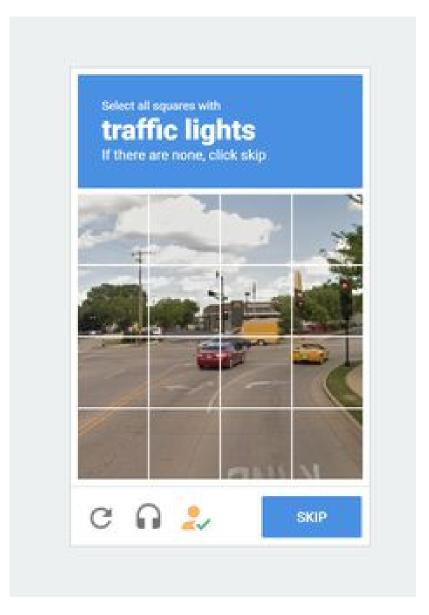
CAPTEHA



(Completely Automated Public Turing test to tell Computers and Humans Apart: test de Turing público y automático para distinguir a los ordenadores de los humanos)

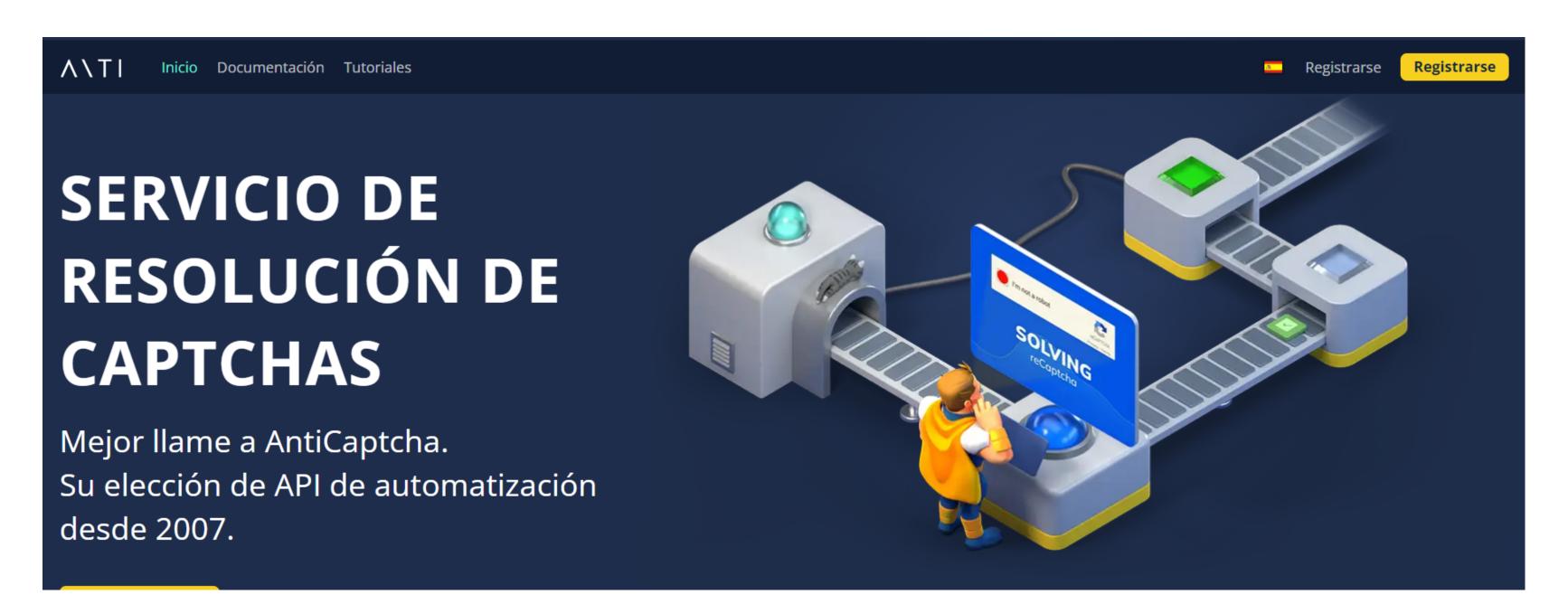






ANTI CAPTCHA





ANTI CAPTCHA



Tipos de tarea soportadas

<u>ImageToTextTask</u>

Resuelve los clásicos captchas de imagen

arch



RecaptchaV2Task

Rompecabezas Recaptcha de Google con un proxy

RecaptchaV2TaskProxyless

Rompecabezas de Google Recaptcha sin proxies



I'm not a robot reCAPTCHA

HCaptchaTask

hCaptcha con un proxy

<u>HCaptchaTaskProxyless</u>

hCaptcha sin proxies









RecaptchaV3TaskProxyless

Google Recaptcha v.3



RecaptchaV2EnterpriseTask

Recaptcha V2 Enterprise de Google con un proxy

RecaptchaV2EnterpriseTaskProxyless

Recaptcha V2 Enterprise de Google sin proxies



I'm not a robot



Recaptcha V3 Enterprise

Recaptcha Enterprise v.3



ANTI CAPTCHA



Ejemplo de solicitud

```
Python Javascript PHP Java C#
                                  bash
//npm install @antiadmin/anticaptchaofficial
//https://github.com/anti-captcha/anticaptcha-npm
const ac = require("@antiadmin/anticaptchaofficial");
const fs = require('fs');
const captcha = fs.readFileSync('captcha.png', { encoding: 'base64' });
ac.setAPIKey('YOUR_API_KEY_HERE');
//Specify softId to earn 10% commission with your app.
//Get your softId here: https://anti-captcha.com/clients/tools/devcenter
ac.setSoftId(0);
ac.solveImage(captcha, true)
    .then(text => console.log('captcha text: '+text))
    .catch(error => console.log('test received error '+error));
```

LIRRERÍAS









PAGINAS PARA PRAETICAR



https://the-internet.herokuapp.com

http://books.toscrape.com/catalogue/category/books/fiction 10/index.html