

# Scraping y parsing de webs HTML

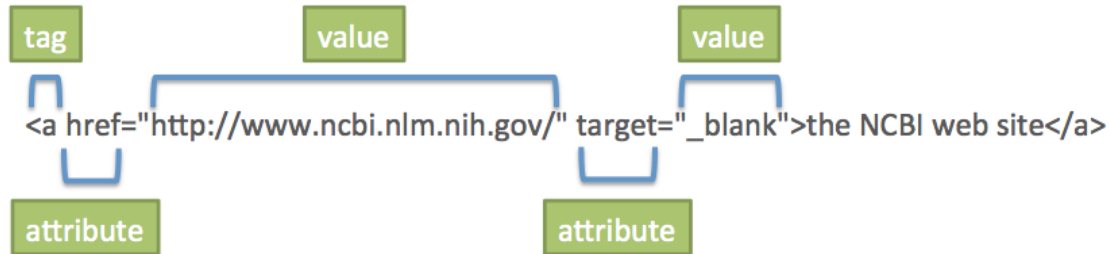


# Web scraping de webs HTML



# Análisis de HTML mediante BeautifulSoup

- **BeautifulSoup** es una librería de Python que permite analizar documentos HTML y extraer datos de ellos, compensando imperfecciones que puedan existir.
- Por ejemplo, permite extraer los atributos href de las etiquetas de anclaje (a)



# Web scraping de webs HTML

```
import urllib.request

fhand = urllib.request.urlopen('https://openwebinars.net/')
for line in fhand:
    print(line.decode().strip())
```

```
...
<li><a href="/aviso-legal-terminos-condiciones/" rel="nofollow noopener">Términos y condiciones de uso</a></li>
<li><a href="/aviso-legal-politica-privacidad/" rel="nofollow noopener">Política de privacidad</a></li>
...
```

CONTACTA

TÉRMINOS Y CONDICIONES DE USO

BLOG

POLÍTICA DE PRIVACIDAD

**¡GRACIAS!**

¿Preguntas?