# **Páginas fáciles para practicar web scraping con find / find\_all**

| **Nº** | **Página** | **Qué extraer** | **Instrucciones con .find / .find\_all** |
| --- | --- | --- | --- |
| 1 | Books to Scrape | Títulos y precios de los libros | Haz soup.find\_all("article", class\_="product\_pod"). Dentro de cada article, usa find("h3").find("a")["title"] para el título y find("p", class\_="price\_color").get\_text(strip=True) para el precio. |
| 2 | Quotes to Scrape | Citas y autores | Haz soup.find\_all("span", class\_="text") para las citas y soup.find\_all("small", class\_="author") para los autores. |
| 3 | Wikipedia – Países por población | Tabla de países y población | Encuentra la primera tabla con soup.find("table", class\_="wikitable"). Luego recorre todas las filas con find\_all("tr") y dentro cada celda con find\_all("td"). |
| 4 | [IMDB Top 250](https://www.imdb.com/chart/top/) | Títulos y año de películas | Encuentra la tabla con soup.find("tbody", class\_="lister-list"). Luego usa find\_all("tr") para las películas, y dentro find("td", class\_="titleColumn").a.get\_text(strip=True) para el título. |
| 5 | ESPN Soccer Scoreboard | Equipos y resultados | Encuentra todos los contenedores de partidos con soup.find\_all("section", class\_="Scoreboard"). Dentro busca equipos con find\_all("span", class\_="ScoreCell\_\_TeamName") y resultados con find\_all("span", class\_="ScoreCell\_\_Score"). |
| 6 | [BBC News – Portada](https://www.bbc.com/news) | Titulares principales | Busca todos los titulares con soup.find\_all("h3"). Filtra los que tienen texto (ej: h3.get\_text(strip=True)). |
| 7 | Wikipedia – Nobel de Física | Tabla de laureados | Encuentra la primera tabla con soup.find("table", class\_="wikitable"). Luego recorre find\_all("tr") y extrae año y nombre con find\_all("td"). |
| 8 | Wikipedia – Euro | Tabla de monedas vs euro | Igual: soup.find("table", class\_="wikitable") y dentro find\_all("tr"). De cada fila, saca find\_all("td"). |
| 9 | [The Guardian – Tech](https://www.theguardian.com/uk/technology) | Titulares de tecnología | Busca artículos con soup.find\_all("h3"). Dentro de cada uno, si hay un <a>, haz h3.find("a").get\_text(strip=True). |
| 10 | [Python.org – Eventos](https://www.python.org/events/python-events/) | Próximos eventos | Encuentra la lista con soup.find("ul", class\_="list-recent-events"). Dentro recorre find\_all("li"). Extrae li.find("h3").a.get\_text(strip=True) para el nombre y li.find("time").get\_text(strip=True) para la fecha. |