

Prueba técnica de Web Scraping en Python.

Descripción:

Crear un script en Python 3.10 que, dada la URL

<https://www.larodan.com/products/category/monounsaturated-fa/>, realice dos operaciones: Lister y Crawler.

- El Lister debe obtener las URLs de productos desde la página proporcionada (tener en cuenta la paginación).
- El Crawler debe recopilar datos específicos de cada producto.
- El resultado final será un JSON con la información solicitada, es decir, un listado de productos con sus datos.

Requisitos:

- No se debe usar ninguna librería externa, excepto para parsear HTML, convertir las imágenes y parsear el PDF. Las librerías estándar se pueden utilizar sin restricciones.
- Crear un archivo `requirements.txt` con las versiones exactas de las librerías utilizadas.
- Implementar el parámetro `-c` en el comando de ejecución del script para permitir el lanzamiento de múltiples Crawlers de forma paralela. Ejemplo: `-c 5` lanzaría 5 Crawlers simultáneos.
- El proceso de obtención de URLs (Lister) no debe ser paralelizado, pero el proceso de Crawler sí.

Opcional:

- README.
- Parsear el PDF para obtener alguna información adicional de este (el `UN Number` es especialmente preferible, está en la sección 14.1, pero puede no estar si el producto no es peligroso).
- Convertir la imagen en una miniatura formato PNG, y comprimir si se desea. Guardarla en un directorio imágenes (crear el directorio si no existe) con nombre que se haga referencia desde el delivery.

Ejemplo del JSON final:

```
[
  {
    "id": "0590S",
    "name": "Bergaptol",
    "CAS": "486-60-2",
    "structure": "C<sub>11</sub>H<sub>6</sub>O<sub>4</sub>",
    "smiles": "OC1=C2C=CC(=O)OC2=CC2=C1C=CO2",
    "description": "Bergaptol description",
    "molecular_weight": "202.17",
    "url": "https://domain.com/product_page.html",
    "image_path": "products/img/0590S.png",
    "img": "https://domain.com/product_page/main_image.png",
    "pdf_msds": "https://domain.com/product_page/sds_document.pdf",
    "synonyms": [
      "4-Hydroxypsoralen",
      "5-Hydroxypsoralen"
    ],
    "packaging": {
      "10mg": 139.00,
      "25mg": 242.50
    }
  },
  {
    "id": "0823M",
    "...": "..."
  }
]
```

Si tienes sugerencias para mejorar la propuesta original, no dudes en implementarlas. En ese caso, es importante que nos envíes la documentación pertinente, explicando los cambios realizados y las razones detrás de estos.