

## WebScrapping – Documentación:

### Integrantes:

- Henry David Martínez Villarraga – 160004125
- Javier Andrés Charry Coronado – 160004108
- Rodrigo Andrés Cruz Morales – 160003810
- Andrés Alejandro Duque Tauta – 160003812

### 1. Listado de requerimientos:

Analizador de: Archivos HTML - WEB SCRAPING  
Tamaño en KBytes

Cantidad de líneas

Tamaño de líneas

Ocurrencia de palabras

Cantidad de enlaces

Listado de enlaces

a otro HTML

a pdf

a otros?

Listado de enlaces: ACTIVOS O ROTOS

Descargar imágenes

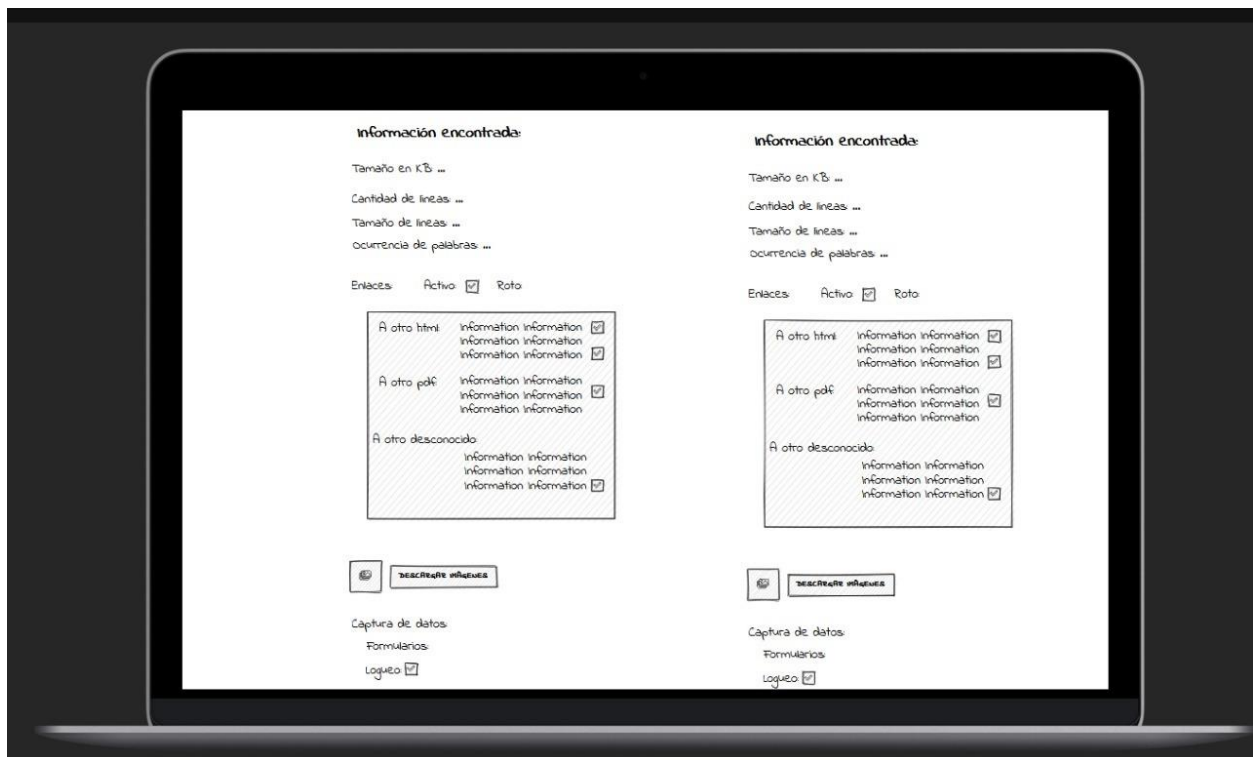
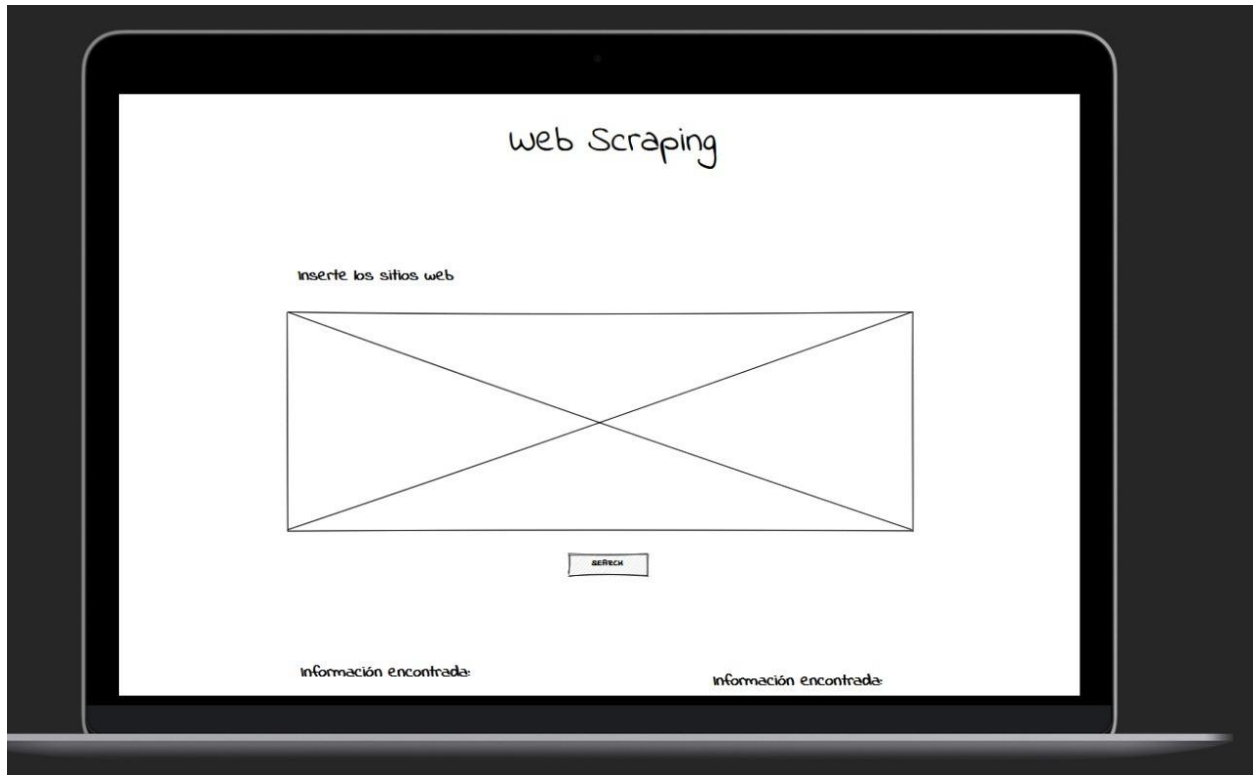
Detectar sitios con captura de datos a través de formularios.

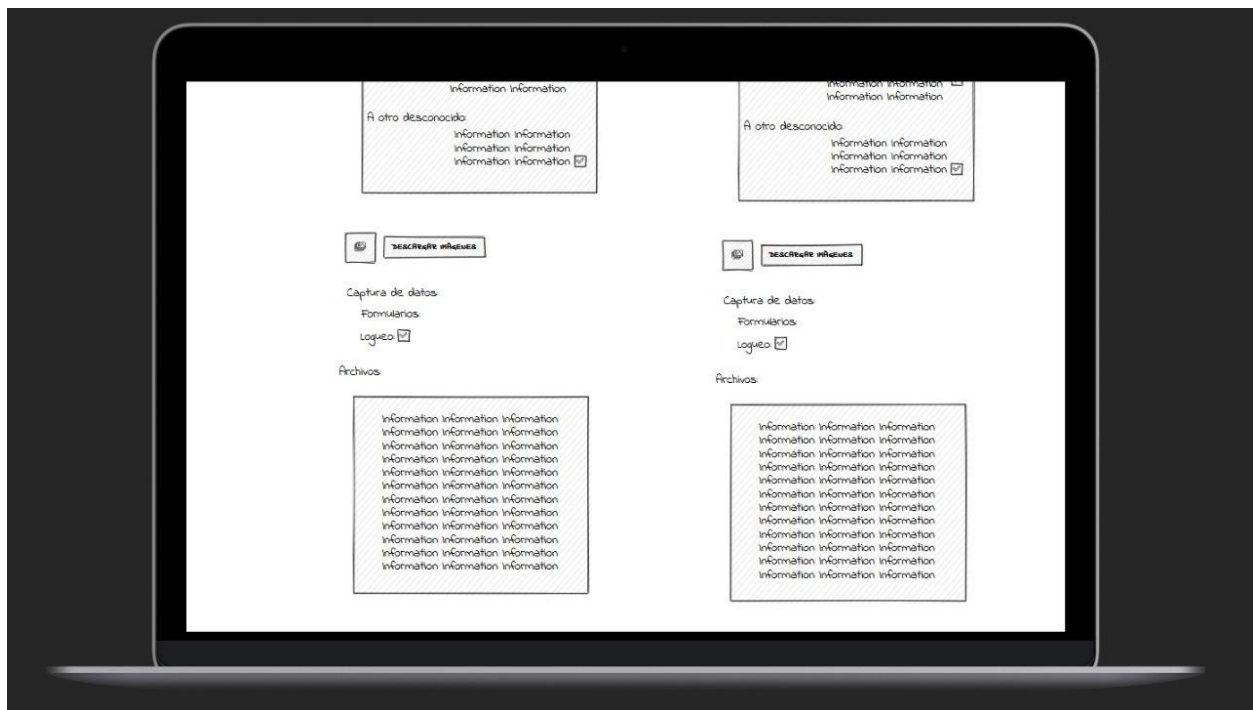
Detectar sitios con logueo.

Listado de archivos.

Todos los proyectos deben implementar procesamiento concurrente: con procesos o hilos.

## 2. Mockups:





### 3. Diagrama de clases UML:

