Web Scraping

Dr. Paul Escapil-Inchauspé Universidad Adolfo Ibáñez, Chile

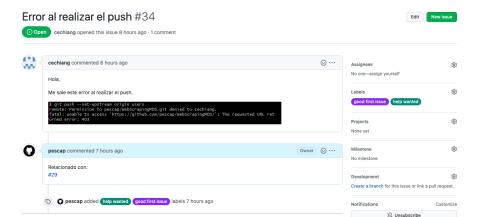
> Proyecto 0 24 de agosto de 2022

Esquema

Presentación del proyecto

2/12

Creación de Issue (state=open)



Resolución de Issue (state=closed)

$\mathsf{Regla} = \mathsf{Closed} + \mathsf{Assigned}$



4 / 12

Resumen

- state: {open, closed}
- author: usuario del author
- assignees: usuarios asignados

5/12

¿Automatizar?

- \bullet Si todos los alumnos crean issues: 3 \times 30 issues que monitorear
- Más trabajo todavía con los issues cerrados
- Más justo (sólo assignees por el profesor y/o ayudante)
- Más motivación

Objetivo final: Evaluación en tiempo real de la nota de los alumnos

1-Pseudo-code:

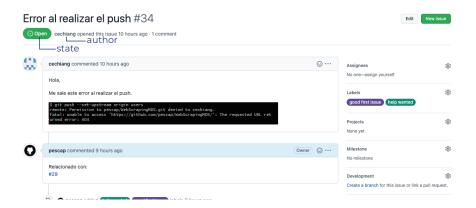
for cada periodo: for cada issue: **buscar: state, author, assignees** agregarlo a una fila de DataFrame

- 2-Realizar manualmente la operación para algunos casos
- 3-Programar la operación "buscar"
- 4-Escalar a todos los issues
- 5-Escalar a todos los periodos
- 6-Desplegar y monitorear

Elección de tecnología

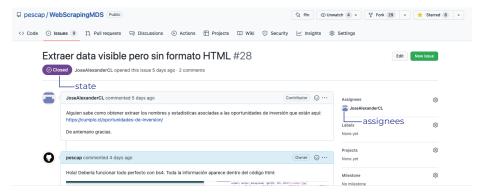
- No hay API
- Código HTML clásico
- No hay acciones o código dinámico
- No parece haber otras restricciones de conexión
- No se necesita login

Inspect: open



Web Scraping

Inspect: closed



T. 1	1067.	46		
	106]:	ат		
	issue	author	state	assignees
0		pescap	open	No one assigned
		pescap	closed	pescap
		pescap	closed	caspoerer
		pescap	closed	No one assigned
		GeorgeAdrock	open	GeorgeAdrock
		pescap	closed	No one assigned
6	15	DenisOsses	closed	No one assigned
		CarlosZuritaEstolaza	open	CarlosZuritaEstolaza
8	20	CarlosZuritaEstolaza	open	No one assigned
	21	gavalenz	open	No one assigned
10	24	shorokein	open	No one assigned
	25	josemoragonzalez	closed	pescap
12	25	josemoragonzalez	closed	josemoragonzalez
13	27	pescap	open	No one assigned
14	28	JoseAlexanderCL	closed	JoseAlexanderCL
15	29	iggypoprl	open	No one assigned
16	34	cechiang	open	No one assigned

Número de columnas: 4.

Estimación número de filas: $38 \times 6 = 228$.

Orden de magnitud esperado 1000 filas aproximadamente.

