

## PRACTICA 2 - MONKEY SCRAPER

Para hacer uso de la aplicación Monkey Scraper debe ingresar la siguiente información

- 1) Cantidad de monos buscadores

```
----- PRACTICA 2 - G12 SOPES2 -----  
----- MONKEY SCRAPER CLI -----  
  
- Ingrese cantidad de monos buscadores  
SOPES2 >> 
```

- 2) El tamaño de la cola

```
----- PRACTICA 2 - G12 SOPES2 -----  
----- MONKEY SCRAPER CLI -----  
  
- Ingrese cantidad de monos buscadores  
SOPES2 >> 5  
  
- Ingrese tamaño de la cola  
SOPES2 >> 30
```

- 3) El valor Nr

```
----- PRACTICA 2 - G12 SOPES2 -----  
----- MONKEY SCRAPER CLI -----  
  
- Ingrese cantidad de monos buscadores  
SOPES2 >> 5  
  
- Ingrese tamaño de la cola  
SOPES2 >> 30  
  
- Ingrese valor de Nr  
SOPES2 >> 3
```

4) La url de búsqueda inicial

```
----- PRACTICA 2 - G12 SOPES2 -----  
----- MONKEY SCRAPER CLI -----  
  
- Ingrese cantidad de monos buscadores  
SOPES2 >> 5  
  
- Ingrese tamaño de la cola  
SOPES2 >> 30  
  
- Ingrese valor de Nr  
SOPES2 >> 3  
  
- Ingrese búsqueda inicial  
SOPES2 >> https://es.wikipedia.org/wiki/James\_Franco
```

5) Nombre del archivo donde se almacenará el resultado

```
----- PRACTICA 2 - G12 SOPES2 -----  
----- MONKEY SCRAPER CLI -----  
  
- Ingrese cantidad de monos buscadores  
SOPES2 >> 5  
  
- Ingrese tamaño de la cola  
SOPES2 >> 30  
  
- Ingrese valor de Nr  
SOPES2 >> 3  
  
- Ingrese búsqueda inicial  
SOPES2 >> https://es.wikipedia.org/wiki/James\_Franco  
  
- Ingrese nombre del archivo donde se almacenará el resultado.  
SOPES2 >> result
```

Mostrará al usuario un resumen de sus configuraciones antes de iniciar con el scrapping

```
Resumen de configuración:
Cantidad de monos -> 5
Tamaño de la cola -> 30
Valor Nr -> 3
URL inicial -> https://es.wikipedia.org/wiki/James\_Franco
Nombre del archivo -> result
Cargando configuración:#####
Configuración cargada correctamente. Presione ENTER para iniciar el scrapping. :D

Ejecutando Monkey Scraper...
```

Los monos deben realizar el siguiente trabajo con cada artículo que leen

1. Contar la cantidad de palabras dentro del artículo de wikipedia.
2. Contar la cantidad de enlaces a otros artículos de wikipedia.
3. Obtener el sha-256 del contenido
4. Para la primera búsqueda se obtienen las n referencias más mencionadas y se agregan a la cola de espera haciendo que  $Nr = Nr - 1$  para esa nueva búsqueda, para que otro mono buscador las analice.
5. Si la cola está llena las referencias son descartadas.
6. El mono debe colocar su id al resultado de la búsqueda y guardarlo en el archivo.
7. El mono debe descansar un tiempo proporcional al número de palabras leídas antes de empezar otra búsqueda.

En la consola se podrá visualizar el recorrido de los monos.

```
El mono monkey_04 esta descansando
[{96f5cd4a97765777e8b01410fc103cbf40245214 https://es.wikipedia.org/wiki/Alfonso\_X\_de\_Castilla 0}]
El mono monkey_02 ya esta disponible
El mono monkey_02 esta descansando
[]
El mono monkey_02 ya esta disponible
El mono monkey_02 esta descansando
El mono monkey_04 ya esta disponible
El mono monkey_03 ya esta disponible
El mono monkey_05 ya esta disponible
El mono monkey_02 ya esta disponible
El mono monkey_01 ya esta disponible
Scraping terminado. Presione ENTER para salir. :D
```

## Archivo Resultante

```
[{
  "Origen": "0",
  "Palabras": 2549,
  "Enlaces": 137,
  "SHA": "c660c93f2375b8d1bd6d08c0a92c36ab1e0a1277",
  "URL": "https://es.wikipedia.org/wiki/James_Franco",
  "Mono": "monkey_01"
},
{
  "Origen": "c660c93f2375b8d1bd6d08c0a92c36ab1e0a1277",
  "Palabras": 1904,
  "Enlaces": 33,
  "SHA": "9c9eeee28a028c32a667c8feade97433b9775627",
  "URL": "https://es.wikipedia.org/wiki/Palo_Alto",
  "Mono": "monkey_02"
},
{
  "Origen": "c660c93f2375b8d1bd6d08c0a92c36ab1e0a1277",
  "Palabras": 2161,
  "Enlaces": 197,
  "SHA": "467dfcc958b23786b813bdb72911d812273a3100",
  "URL": "https://es.wikipedia.org/wiki/California",
  "Mono": "monkey_03"
},
{
  "Origen": "c660c93f2375b8d1bd6d08c0a92c36ab1e0a1277",
  "Palabras": 457,
  "Enlaces": 23,
  "SHA": "4698478b9ce32b4518ecb790bc2a1d98e4ffec7a",
  "URL": "https://es.wikipedia.org/wiki/Actor",
  "Mono": "monkey_04"
},
{
  "Origen": "9c9eeee28a028c32a667c8feade97433b9775627",
  "Palabras": 926,
  "Enlaces": 48,
  "SHA": "27eb8fbc576d552b0c4f089ba2f03c1c83e877d",
```

```
"URL": "https://es.wikipedia.org/wiki/Ciudades_de_Estados_Unidos",
"Mono": "monkey_05"
},
{
  "Origen": "9c9eeee28a028c32a667c8feade97433b9775627",
  "Palabras": 1034,
  "Enlaces": 80,
  "SHA": "035e557784c5019d26f39ad6b9fb4a2441585238",
  "URL": "https://es.wikipedia.org/wiki/Condado_de_Santa_Clara_(California)",
  "Mono": "monkey_02"
},
{
  "Origen": "467dfcc958b23786b813bdb72911d812273a3100",
  "Palabras": 1796,
  "Enlaces": 83,
  "SHA": "2aa50f5dd42323c033b52b35f940727b049ea671",
  "URL": "https://es.wikipedia.org/wiki/Estado_de_los_Estados_Unidos",
  "Mono": "monkey_01"
},
{
  "Origen": "467dfcc958b23786b813bdb72911d812273a3100",
  "Palabras": 4946,
  "Enlaces": 322,
  "SHA": "3318d2062ab84ee78cf04e941a0249f6f6ec24af",
  "URL": "https://es.wikipedia.org/wiki/Washington_D._C.",
  "Mono": "monkey_03"
},
{
  "Origen": "4698478b9ce32b4518ecb790bc2a1d98e4ffec7a",
  "Palabras": 2987,
  "Enlaces": 91,
  "SHA": "d7cdc3bb015d1a24edd18e2efafa26ef333d275b",
  "URL": "https://es.wikipedia.org/wiki/Autor",
  "Mono": "monkey_02"
},
{
  "Origen": "4698478b9ce32b4518ecb790bc2a1d98e4ffec7a",
```

```
"Palabras": 1467,
"Enlaces": 89,
"SHA": "96f5cd4a97765777e8b01410fc103cbf40245214",
"URL": "https://es.wikipedia.org/wiki/T%C3%ADteres",
"Mono": "monkey_04"
},
{
  "Origen": "27eb8fbc576d552b0c4f089ba2f03c1c83e877d",
  "Palabras": 13320,
  "Enlaces": 1075,
  "SHA": "24355b4053f7952209622ab6f95d66410aaa4262",
  "URL": "https://es.wikipedia.org/wiki/Estados_Unidos",
  "Mono": "monkey_01"
},
{
  "Origen": "035e557784c5019d26f39ad6b9fb4a2441585238",
  "Palabras": 3757,
  "Enlaces": 281,
  "SHA": "e5e05e1e5b65d765b8463155552a7a03ab6af7ba",
  "URL": "https://es.wikipedia.org/wiki/Idioma_ingl%C3%A9s",
  "Mono": "monkey_05"
},
{
  "Origen": "2aa50f5dd42323c033b52b35f940727b049ea671",
  "Palabras": 321,
  "Enlaces": 46,
  "SHA": "0d66b72042b664ee54029abae30036bac3e0bbae",
  "URL":
"https://es.wikipedia.org/wiki/Demarcaci%C3%B3n_administrativa",
  "Mono": "monkey_04"
},
{
  "Origen": "3318d2062ab84ee78cf04e941a0249f6f6ec24af",
  "Palabras": 1799,
  "Enlaces": 104,
  "SHA": "54dac58b9c90d45b5914b23e2749e9445a1b55f5",
  "URL":
"https://es.wikipedia.org/wiki/Alfabeto_Fon%C3%A9tico_Internacional",
  "Mono": "monkey_04"
}
```

```
}  
,  
{  
  "Origen": "d7cdc3bb015d1a24edd18e2efafa26ef333d275b",  
  "Palabras": 650,  
  "Enlaces": 58,  
  "SHA": "b89b22f2c5f8899226be638094be1c3ab5cef4f6",  
  "URL": "https://es.wikipedia.org/wiki/Obra_art%C3%ADstica",  
  "Mono": "monkey_02"  
}  
,  
{  
  "Origen": "96f5cd4a97765777e8b01410fc103cbf40245214",  
  "Palabras": 6979,  
  "Enlaces": 354,  
  "SHA": "3710eb91a66c25f4b25a08f8e9214ddded826dc2",  
  "URL": "https://es.wikipedia.org/wiki/Alfonso_X_de_Castilla",  
  "Mono": "monkey_02"  
}  
]
```