

Trabajo Práctico N°1

Integrantes

- Milla, Andrés - 14934/6
- Mandarinino, Leonel - 15379/5

Notas

- Este archivo representa el archivo **src.txt** mencionado en el enunciado.
- Las notas se encuentran en el directorio **doc**. Se recomienda leer el PDF para una mejor experiencia :)
- En el directorio **data** se encuentran los siguientes archivos:
 - cinemalaplata.json: Es la información recolectada de Cinema La Plata.
 - cinopolis.json: Es la información recolectada de Cinépolis.
 - movies.json: Es la mezcla de ambos scrapers, se utilizó una estrategia de mezcla para evitar inconsistencias y unificar la información (Para más información leer el archivo **doc/notas.pdf**)

Requisitos

- *Python3.8* o versiones posteriores.
- *Chromedriver* : <https://chromedriver.chromium.org/downloads>
- Se recomienda un entorno virtual como *virtualenv* para instalar las dependencias

Ejecución

1. Copiar el binario de *Chromedriver* en el directorio **bin**
2. Crear y activar el entorno virtual:

```
virtualenv -p python3 venv
source venv/bin/activate
```

3. Instalar las dependencias:

```
pip3 install -r requirements.txt
```

4. Ir al directorio fuente:

```
cd src
```

5. Ejecutar el script:

```
python3 main.py
```

• Opciones de ejecución

`--verbose` : Permite visualizar todos los logs.

`--offline` : Ejecuta solo el merge, es útil para no scrapear de nuevo.