

# Scrapings

A continuación exponemos el **Payload (Schema)** necesario y **obligatorio** en todos los scripts de extracción desarrollados en la **Scraping**.

```
payload_contenidos = {
  "PlatformCode": "str", #Obligatorio
  "Id": "str", #Obligatorio
  "Seasons": [ #Unicamente para series
    {
      "Id": "str", #Importante
      "Synopsis": "str", #Importante
      "Title": "str", #Importante, E.J. The Wallking Dead:
Season 1
      "Deeplink": "str", #Importante
      "Number": "int", #Importante
      "Year": "int", #Importante
      "Image": "list",
      "Directors": "list", #Importante
      "Cast": "list", #Importante
      "Episodes": "int" #Importante
      "IsOriginal": "bool" packages
    },
    ...
  ],
  "Crew": [ #Importante
    {
      "Role": str,
      "Name": str
    },
    ...
  ],
  "Title": "str", #Obligatorio
  "CleanTitle": "_replace(str)", #Obligatorio
  "OriginalTitle": "str",
  "Type": "str", #Obligatorio #movie o serie
  "Year": "int", #Important! 1870 a año actual
  "Duration": "int", en minutos
  "ExternalIds": "list", *
  "Deeplinks": {
    "Web": "str", #Obligatorio
    "Android": "str",
    "iOS": "str",
  },
  "Synopsis": "str",
  "Image": "list",
  "Rating": "str", #Important!
```

```

"Provider":      "list",
"Genres":        "list",      #Important!
"Cast":          "list",      #Important!
"Directors":      "list",      #Important!
"Availability":   "str",       #Important!
"Download":       "bool",
"IsOriginal":     "bool",      #Important!
"IsAdult":        "bool",      #Important!
"IsBranded":      "bool",      #Important!    (ver link explicativo)
"Packages":       "list",      #Obligatorio
"Country":        "list",
"Timestamp":      "str", #Obligatorio
"CreatedAt":      "str", #Obligatorio
}

```

```

payload_episodios = {
  "PlatformCode": "str", #Obligatorio
  "Id":            "str", #Obligatorio
  "ParentId":      "str", #Obligatorio #Unicamente en Episodios
  "ParentTitle":   "str", #Unicamente en Episodios
  "Episode":       "int", #Obligatorio #Unicamente en Episodios
  "Season":        "int", #Obligatorio #Unicamente en Episodios
  "Crew":           [ #Importante
    {
      "Role": str,
      "Name": str
    },
    ...
  ],
  "Title":          "str", #Obligatorio
  "OriginalTitle":  "str",
  "Year":           "int",   #Important!
  "Duration":       "int",
  "ExternalIds":    "list", *
  "Deeplinks": {
    "Web":          "str",      #Obligatorio
    "Android":      "str",
    "iOS":          "str",
  },
  "Synopsis":       "str",
  "Image":          "list",
  "Rating":         "str",      #Important!
  "Provider":       "list",
  "Genres":         "list",      #Important!
  "Cast":           "list",      #Important!
  "Directors":       "list",      #Important!
  "Availability":   "str",      #Important!
  "Download":       "bool",
  "IsOriginal":     "bool",      #Important!
  "IsAdult":        "bool",      #Important!
  "IsBranded":      "bool",      #Important!    (ver link explicativo)
}

```

```
"Packages":      "list",      #Obligatorio
"Country":       "list",
"Timestamp":     "str", #Obligatorio
"CreatedAt":     "str", #Obligatorio
}
```

## Variables

---

**PlatformCode:** Código único de la plataforma, con formato [Código País].[Nombre Plataforma]. [Ver más](#)

**Id:** Identificación única del contenido, tal como viene de la plataforma. En caso de no conseguirla, se genera.

**Title:** Título del contenido, tal como viene de la plataforma.

**CleanTitle:** Título del contenido, que pasa por el método `_replace()` para limpiar el título.

**OriginalTitle:** Título del contenido, traducido a su idioma original.

**Type:** Tipo de contenido. Siempre tiene que ser *movie* o *serie*.

**Year:** Año de producción del contenido.

**Duration:** Duración del contenido. Siempre en minutos.

**ExternalIds:** Id de BBDDs externas (imdb, tvdb, tmdb, eidr, etc) si se puede identificar en la plataforma.

**Ejemplo:**

```
[
  {
    "Provider": "IMDb",
    "Id": "tt12098392"
  }
]
```

**Deeplinks:** URLs del contenido.

**Synopsis:** Descripción del contenido.

**Image:** Poster del contenido.

**Rating:** Calificación de edad del contenido. **Ejemplos:** PG13, TV-MA, FSK 18, etc.

**Provider:** Productora del contenido.

**Genres:** Géneros del contenido, según aparecen en la plataforma.

**Cast:** Reparto del contenido.

**Crew:** Equipo de producción diferente a director o actores. Por ejemplo productores, escenario, escritores, etc.

*Aclaracion: Si hay dos productores por ejemplo, estos deben venir por separado y no hay que hacer una lista dentro de "Name". Por ejemplo si tenemos dos productores, debe quedar de la siguiente manera:*

```
"Crew": [
  {
    "Role": "Producer",
    "Name": "Pepito"
  },
  {
    "Role": "Producer",
    "Name": "Juan"
  }
]
```

```

    {
      "Role": "Producer",
      "Name": "Juanita"
    },
    ...
  ]

```

**Directors:** Director/es del contenido.

**Availability:** Fecha en la que expira el contenido en la plataforma.

**Download:** Se puede descargar el contenido?

**IsOriginal:** Es un contenido original?

**IsAdult:** Es un contenido para adultos?

**Packages:** Modelo de negocio del contenido. Ejemplo:

```

{"Type": "str"}

```

El campo **Type** muestra como ofrece el contenido la plataforma, ya sea gratis ( **free-vod** ), servicio de cable ( **tv-everywhere** ), suscripción ( **subscription-vod** ) o por pago por contenido ( **transaction-vod** ), o servicio de telefonía/internet ( **validated-vod** )

Para mas info sobre **validated-vod** pueden mirar el siguiente artículo: <https://bb.vision/infografias/nuevo-analisis-planes-comerciales-otts/>

Cuando **Type** es **transaction-vod** se deben utilizar campos adicionales:

```

{
  "BuyPrice": "float",
  "RentPrice": "float",
  "Definition": "str",
  "Currency": "str"
}

```

Donde:

**RentPrice** es el precio de alquiler (usualmente por 48hs),

**BuyPrice** es el precio de compra,

**Definition** es la calidad de video (**SD, HD, 4K, 8K**) [Casos UHD dejar como 4K]

**Currency** es la moneda en la que se hace la transacción (**USD, EUR, ARS**)

Finalmente, un ejemplo de un contenido que se puede alquilar y comprar:

```

[
  {
    "Type" : "transaction-vod",
    "RentPrice" : 4.99,
    "BuyPrice" : 9.99,
    "Definition" : "HD",
    "Currency" : "USD"
  },

```

```
{
  "Type" : "transaction-vod",
  "RentPrice" : 1.99,
  "BuyPrice" : 2.99,
  "Definition" : "SD",

}
]
```

**Country:** Pais de origen del contenido.

**Timestamp:** Fecha y hora en la que se genero el registro.

**CreatedAt:** Fecha de scraping. Formato YYYY-MM-DD.

**Seasons:** Solo incluir en el payload de titanScraping dentro de cada serie. Los cuatro primeros campos son los primordiales, en caso de que se encuentren en la plataforma. El resto de los campos, tienen menor prioridad y solo se deben agregar si son especialmente de la seccion de la temporada en si (no recopilar desde los episodios).

**isBranded:** Documento explicativo en el siguiente link con algunas maneras de traerla

<https://businessbureau0->

[my.sharepoint.com/:w:/g/personal/jm\\_bb\\_vision/ESdmwOadaDFGpKvK9DM2atgBQw7-Isfc6sV5Xb418BbNRA?e=Xn5HOe](https://my.sharepoint.com/:w:/g/personal/jm_bb_vision/ESdmwOadaDFGpKvK9DM2atgBQw7-Isfc6sV5Xb418BbNRA?e=Xn5HOe)

## Cómo hacer un package correctamente

Indicamos cómo hacer un **package correcto**, según los siguientes casos...

**IMPORTANTE:** No hardcodear y buscar algún **"tag html"** o **"key de la API"** para indicar el package correcto del contenido.

Si es posible, podemos validar con un **raise** su modelo de negocio, en caso de que los **packages** cambien a futuro:

```
try:
    access = json_epi["potentialAction"]["actionAccessibilityRequirement"]
["category"]
    if access == "subscription":
        package = {"Type":"subscription-vod"}
    elif access == "nologinrequired":
        package = {"Type":"free-vod"}
    else:
        print(json_epi)
        raise Exception(f"Package type not accounted for in {deeplink_epi}")
except Exception as e:
    print(e)
    continue
```

Básicamente los casos que puede haber son:

- **Caso 1:** La plataforma solo es **subscription-vod**, **free-vod** o **tv-everywhere**.
- **Caso 2:** Caso en las que hay contenidos que son **subscription-vod**, o **tv-everywhere**, pero tiene algunos episodios gratuitos.
- **Caso 3:** Caso en las que hay contenidos **transaccion-vod**.
- **Caso 4:** Caso en las que hay contenidos **transaccion-vod** y **subscription-vod**.
- **Caso 5:** Caso en las que hay **series transaccion-vod** (Alquiler o Compra de temporada). **>CADUCADO**
- **Caso 6:** Caso en las que hay **series transaccion-vod** (Alquiler o Compra de temporada). **NUEVO CASO, USAR ESTE**

Los cuales, iremos comentando a continuación:

**Caso 1:** La plataforma solo es **subscription-vod**, **free-vod** o **tv-everywhere**.

#### En titanScraping

```
{
  "Id": "1abc",
  "Title": "Pepito",
  // Otras "Keys"...
  "Packages": [
    {
      "Type" : "subscription-vod", // Puede ser "free-vod" o "tv-everywhere".
    }
  ],
}
```

#### En titanScrapingEpisodes.

```
{
  "ParentId": "1abc",
  "ParentTitle": "Pepito",
  // Otras "Keys"...
  "Packages": [
    {
      "Type" : "subscription-vod", // Puede ser "free-vod" o "tv-everywhere".
    }
  ],
}
```

**Caso 2:** Caso en las que hay **series** que son **subscription-vod** o **tv-everywhere**, pero tiene algunos episodios gratuitos.

En **titanScraping** (Con el hecho de ya tener un episodio gratuito, el package debe ir de esta manera)

```
{
  "Id": "1abc",
  "Title": "Pepito",
  // Otras "Keys"...
  "Packages": [
    {
      "Type" : "subscription-vod" // Puede ser "tv-everywhere".
    },
    {
      "Type" : "free-vod"
    }
  ],
}
```

En **titanScrapingEpisodes**.

Episodio Gratuito:

```
{
  "ParentId": "1abc",
  "ParentTitle": "Pepito",
  // Otras "Keys"...
  "Packages": [
    {
      "Type" : "free-vod"
    }
  ],
}
```

Episodio Pago:

```
{
  "ParentId": "1abc",
  "ParentTitle": "Pepito",
  // Otras "Keys"...
  "Packages": [
    {
      "Type" : "subscription-vod" // Puede ser "tv-everywhere".
    }
  ],
}
```

**Caso 3:** Caso en las que hay contenidos **transaction-vod**.

**Movies en titanScraping:**

```
{
  "Id": "1abc",
  "Title": "Pepito",
```

```

    "Type": "movie",
    // Otras "Keys"...
    "Packages": [
        {
            "Type" : "transaction-vod",
            "RentPrice" : 4.99, // Puede haber RentPrice solo.
            "BuyPrice" : 9.99, // Puede haber BuyPrice solo.
            "Definition" : "HD", // Puede ser SD, HD o 4K.
            "Currency" : "USD"
        }
    ],
}

```

**IMPORTANTE:** Los packages son listas con diccionarios en cada uno de sus índices. Cada índice, representa a la calidad del contenido, donde pueden ser **SD, HD o 4K**.

**Ejemplo:**

```

{
    "Id": "1abc",
    "Title": "Pepito",
    "Type": "movie",
    // Otras "Keys"...
    "Packages": [
        {
            "Type" : "transaction-vod",
            "RentPrice" : 4.99, // Puede haber RentPrice solo.
            "BuyPrice" : 9.99, // Puede haber BuyPrice solo.
            "Definition" : "HD", // Puede ser SD, HD o 4K.
            "Currency" : "USD"
        },
        {
            "Type" : "transaction-vod",
            "RentPrice" : 3.99, // Puede haber RentPrice solo.
            "BuyPrice" : 8.99, // Puede haber BuyPrice solo.
            "Definition" : "SD", // Puede ser SD, HD o 4K.
            "Currency" : "USD"
        }
    ],
}

```

**Serie en titanScraping:**

```

{
    "Id": "1abc",
    "Title": "Pepito",
    "Type": "serie",
    // Otras "Keys"...
    "Packages": [
        {
            "Type" : "transaction-vod",

```



```

    }
  ],
}

```

### Episodio en titanScrapingEpisodes: (El RentPrice y BuyPrice es el precio POR EPISODIO)

```

{
  "ParentId": "1abc",
  "ParentTitle": "Pepito",
  // Otras "Keys"...
  "Packages": [
    {
      "Type" : "transaction-vod",
      "RentPrice" : 4.99, // Puede haber RentPrice solo cuando rentamos el capítulo.
      "BuyPrice" : 9.99, // Puede haber BuyPrice solo cuando compramos el capítulo.
      "Definition" : "HD", // Puede ser SD, HD o 4K.
      "Currency" : "USD"
    }
  ],
}

```

**IMPORTANTE:** Los packages son listas con diccionarios en cada uno de sus índices. Cada índice, representa a la calidad del contenido, donde pueden ser SD, HD o 4K. Ver ejemplo que hicimos en este caso 3 con las movies.

### Caso 4: Caso en las que hay contenidos transaccion-vod y subscription-vod.

#### Movies en titanScraping:

```

{
  "Id": "1abc",
  "Title": "Pepito",
  "Type": "serie",
  // Otras "Keys"...
  "Packages": [
    {
      "Type" : "subscription-vod",
    },
    {
      "Type" : "transaction-vod",
      "RentPrice" : 4.99, // Puede haber RentPrice solo.
      "BuyPrice" : 9.99, // Puede haber BuyPrice solo.
      "Definition" : "HD", // Puede ser SD, HD o 4K.
      "Currency" : "USD"
    }
  ],
}

```

### Serie en titanScraping:

```
{
  "Id": "1abc",
  "Title": "Pepito",
  "Type": "serie",
  // Otras "Keys"...
  "Packages": [
    {
      "Type" : "subscription-vod",
    },
    {
      "Type" : "transaction-vod",
    }
  ],
}
```

### Episodio en titanScrapingEpisodes:

```
{
  "ParentId": "1abc",
  "ParentTitle": "Pepito",
  // Otras "Keys"...
  "Packages": [
    {
      "Type" : "subscription-vod",
    },
    {
      "Type" : "transaction-vod",
      "RentPrice" : 4.99, // Puede haber RentPrice solo cuando rentamos el capítulo.
      "BuyPrice" : 9.99,  // Puede haber BuyPrice solo cuando compramos el capítulo.
      "Definition" : "HD", // Puede ser SD, HD o 4K.
      "Currency" : "USD"
    }
  ],
}
```

### Caso 5: Caso en las que hay series transaccion-vod (Alquiler o Compra de temporada).->CADUCADO

### Serie en titanScraping:

```
{
  "Id": "1abc",
  "Title": "Pepito",
  "Type": "serie",
  // Otras "Keys"...
  "Packages": [
    {
```

```

        "Type" : "transaction-vod",
    }
],
}

```

### Episodio en titanScrapingEpisodes: (Solo se puede comprar o alquilar la temporada)

```

{
  "ParentId": "1abc",
  "ParentTitle": "Pepito",
  // Otras "Keys"...
  "Packages": [
    {
      "Type" : "transaction-vod",
      "SeasonPrice" : 14.99, // Puede haber SeasonPrice cuando rentamos o compramos
      la temporada.
      "Definition" : "HD", // Puede ser SD, HD o 4K.
      "Currency" : "USD"
    }
  ],
}

```

### Caso 6: Caso en las que hay series transaccion-vod (Alquiler o Compra de temporada). **NUEVO CASO, USAR ESTE**

Este tema surgió, porque hay temporadas que se pueden comprar o rentar, y el dato SeasonPrice no sería el adecuado.

### Serie en titanScraping:

```

{
  "Id": "1abc",
  "Title": "Pepito",
  "Type": "serie",
  // Otras "Keys"...
  "Packages": [
    {
      "Type" : "transaction-vod",
    }
  ],
}

```

### Episodio en titanScrapingEpisodes: (Solo se puede comprar o alquilar la temporada y además, se puede comprar o alquilar un capítulo)

```

{
  "ParentId": "1abc",
  "ParentTitle": "Pepito",

```

```
// Otras "Keys"...
"Packages": [
  {
    "Type" : "transaction-vod",
    "SeasonRentPrice" : 14.99, // Puede haber SeasonRentPrice solo cuando rentamos
    la temporada.
    "SeasonBuyPrice" : 19.99, // Puede haber SeasonBuyPrice solo cuando compramos
    la temporada.
    "RentPrice" : 4.99, // OPCIONAL (Si se alquila por capítulo).
    "BuyPrice" : 9.99, // OPCIONAL. (Si se vende por capítulo).
    "Definition" : "HD", // Puede ser SD, HD o 4K.
    "Currency" : "USD"
  }
],
}
```

#### **Draft: Las series se pueden...**

- Comprar/Alquilar por capítulo
- Comprar/Alquilar por temporada
- Comprar/Alquilar un rango de temporadas.
- Comprar/Alquilar toda la serie completa, con todas las temporadas.