# CV Documentation

Una vez finalizado un evento dispondremos de un SS con los datos de cada uno de los participantes. El formato de cada uno de ellos debe ser el siguiente:

**Preprocesar el SS**

Probablemente con el nuevo formulario que tenemos, el 95% de la gente lo haya hecho bien y tenemos unos campos del SS consistentes peeeero hay un 5% que siempre la caga y a lo mejor algún día varía el SS o algo así que el primer paso es echarle un vistazo al SS.

El SS debe tener los siguientes campos mínimos, si tiene más y no pasa nada pero probablemente no se tengan en cuenta todavía:

* full\_name: Nombre de la persona. Formato: Manuel Montoya Catalá.
  + Hay gente gilipollas que escribe sólo su primer nombre ó escribe Montoya Catalá, Manuel y demás. Normalizarlo en la medida de lo posible.
* birthday: Fecha de nacimiento. Formato: DD/MM/AAAA
  + No pasa nada si está en otros formatos como D/M/AA ó usando “ - “ pero si lo han escrito con palabras hay que cambiarlo.
* phone: Número de teléfono. Formato XXXXXXXXX. Todo junto.
* email: Mail de la peña. Este lo suelen poner siempre bien.
* origin: Evento del que proviene. Estos es el evento del que proviene el CV, como por ejemplo “EBEC14”.
  + Si no existe esta columna, créala y ponle a todos el hashtag del evento.
  + Si el evento es nuevo, muy probablemente habrá que crearlo también en la BBDD, lo veremos más adelante.
* current\_studies: Es el grado o Máster que están cursando. Sólo debería pemitírseles poner alguno de los que de verdad existen de entre una lista que les demos pero de no ser así tendréis que corregir los que estén mal.
* year\_start: Año en el que empezaron a estudiar en la universidad. Se utiliza como estimación de en qué año están.

**El orden en el que estén dispuestos los campos no importa !!**

**Crear Evento en la BBDD**

Cada CV tiene asociado el evento en el que fue recogido, para asociarle una fecha temporal. Algunos de los eventos son:

["SC14", 2014]

["EBEC14", 2014]

["TTD15", 2015]

["EBEC15", 2015]

Cada evento tiene una fecha temporal asociada también.

Si el evento del que vienen los CV es nuevo, deberemos primero crearlo, para ello utilizamos el fichero add\_new\_event.py donde indicamos el nombre del evento y su tiempo mediante las variables:

* name = "SC99"
* time = 2099

**Añadir los CV a la BBDD**

Esto consiste en sólo añadir los datos del SS a la BBDD total, que se encargará de normalizarlo, detectar duplicados, actualizar los que deba y demás. Para ello primeramente necesitamos convertir el SS posprocesado a formato .csv con:

* Coma por delimitador “,”
* Nada como dilimitador de texto. A veces te dan a elegir “, borralo please.

Después sólo tenemos que ejecutar el archivo add\_csv\_to\_MasterTable.py indicando la dirección del CSV en la variable:

* tables\_dir = ["./data/CV\_Book.csv"]

**Descargar y nombrar los pdf**

Para poder descargar los “pdf” necesitaremos las url de los archivos y el nombre y cumpleaños de la gente. Lo ideal sería que cogierais las columnas de “full\_name”, “birthday” y “pdfurl”, las copiaseis en un nuevo SS y lo guardaseis en formato .csv con:

* Coma por delimitador “,”
* Nada como dilimitadore de texto. A veces te dan a elegir “, borralo please.

Ahora nuestro archivo XXXXX.csv está listo para ser leído y procesado. Para ello hay que utilizar el archivo **download\_pdf\_events.py**, donde tendremos que especificar:

file\_dir = "./data/TTDpdfs.csv" con la dirección relativa de nuestro fichero.

Todo esto sería fantástico si no fuera por ciertas anomalías:

1. Alguna gente sube su CV en formato word…
2. Alguna gente pone comas “,” en el nombre del archivo
3. Por algún motivo algunos .pdf no se

La gente que lo suba en otro formato que no sea “pdf” hay que sacrificarlos, pero de no poder ser así, lo que pasará es que se descargará el fichero con el nombre adecuado de la persona. Lo único que hay que hacer es cambiarle la extension con Microsoft Office o lo que toque, el nombre se mantiene el mismo.

Los que pongan “,” en el nombre del archivo, estamos jodidos, tendréis que descargarlo a mano.

Los que por alguna razón no se puedan descargar tendréis que hacerlo manualmente y nombrarlos. No temáis, se habrá creado en la carpeta principal un fichero llamado:

./NotDownloadedTable.csv"

Este fichero contiene la url y el nombre que debeis dar a todos los archivos que no pudieron ser descargados.

Cómo te quedas ?

**Añadir los pdfs a la carpeta principal**

Una vez tengas en una carpeta bonita todos los pdf, ya estás preparado para moverlos a la carpeta “all\_pdfs” a que re reunan con el resto, pero por si acaso aparecen duplicados, se hace mediante un fichero que mira el evento de cada uno de ellos y decide si moverlo o no.

Este archivo es **add\_pdfs\_to\_all\_pdfs.py**, al cual tendremos que especificar:

* Carpeta donde están los nuevos pdf: folder\_in = "./pdfs\_down/"
* Evento al que pertencen los pdf: origin\_table = "TTD15"

El archivo se encargará del resto. Si por casualidad había mal nombrado algún archivo:

* Puede darse por violación de la conveniencia “-”

Entonces también te lo dice y tal.

Otras funcionalidades:

**Nombrado automatico de archivos:**

Si por desgracia tenemos una carpeta llena de archivos con el nombre que sean y tenemos que nombrarlos.

El archivo XXXX es capaz de leer los pdfs, buscar los nombres de las pesonas en 2 formatos:

* Manuel Montoya Catalá
* Montoya Catalá, Manuel

Si algún gilipollitas ha hecho algo cómo

Nombre: Manuel

Apellidos: Montoya Catalá

No va a funcionar.

Esta técnica puede ahorrarnos un bonito tiempo de trabajo en tal caso. Los que no consiguiese cambiar pues habría que ponerles el nombre a mano.