

Manual de Usuario ArchivoMex Xtractor – Censos Hospitales.

El procesamiento de los censos de hospitales tiene el siguiente flujo:

1. Cada página del PDF consta de 2 páginas del censo de hospitales. Para segmentar cada página, se recorta la imagen original en 2 partes.
2. Dentro de cada página del censo de hospitales se realiza un proceso de segmentación por rectángulos, este proceso puede generar “ruido” o imágenes que no corresponden a una tabla.
3. Al procesar cada tabla de manera individual, se realiza el mismo proceso que en los anuarios estadísticos, la segmentación por columnas y filas, con la diferencia de que estas tablas presentan un grado de rotación que debe de corregirse.
4. Cada celda se procesa con el OCR para obtener el texto dentro de la imagen.

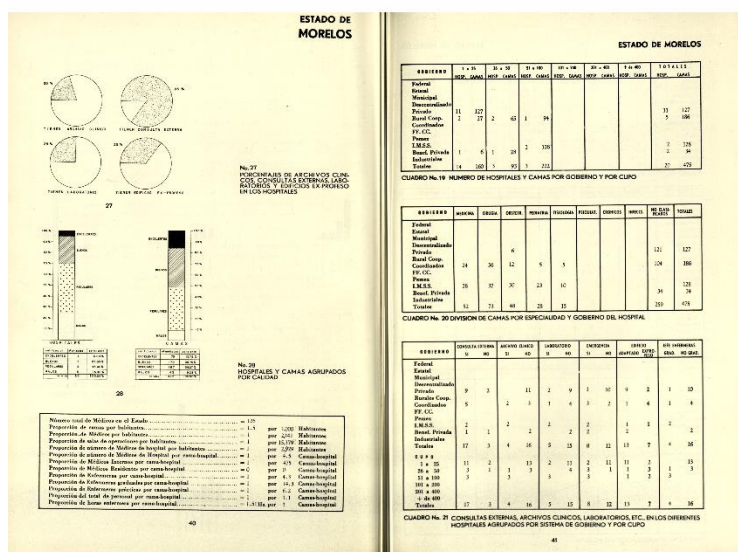
Para utilizar la herramienta, es necesario seguir las siguientes instrucciones:

1. Ejecutar el script y esperar a que la interfaz aparezca en la pantalla.
2. Seleccionar el PDF, el número de página del PDF, el número de página del censo y el ángulo o rotación que tiene la hoja del PDF (-90 si está girada a la izquierda, 90 si está girada a la derecha).
3. Esperar a que se generen las imágenes de cada tabla.
4. Posteriormente, del lado derecho de la interfaz, seleccionar la tabla que se desea procesar.
5. Indicar que columnas son dobles y que columnas tienen texto (enumeradas desde 1 hasta n).
6. Seleccionar la casilla de “Extracción 2” para mejores resultados.
7. Generar archivo de Excel.

Consultar anexo para información adicional.

Anexo.

Ejemplo de input en la interfaz.



En este caso:

Página (1,2): 2

Rotar: 0

GOBIERNO	1 a 25		26 a 50		51 a 100		101 a 200		201 a 400		401 a 600		TOTALES	
	HOSP.	CAMAS	HOSP.	CAMAS	HOSP.	CAMAS	HOSP.	CAMAS	HOSP.	CAMAS	HOSP.	CAMAS	HOSP.	CAMAS
Federal														
Estatal														
Municipal														
Descentralizado														
Privado	11	127											11	127
Rural Coop.	2	27	2	65	1	94							5	186
Coordinados														
FF. CC.														
Pemex														
I.M.S.S.					2	128							2	128
Benef. Privada	1	6	1	28									2	34
Industriales														
Totales	14	160	3	93	3	222							20	475

Columnas dobles: 2,3,4,5,6,7,8

*Una columna doble es aquella que en el encabezado tiene dos categorías, en este caso, "hospitales" y "camas"

Columnas texto: 1

