

# Prosopografía de la Revolución Mexicana: actualización de la obra de Francoise Xavier Guerra

Diego Montesinos<sup>1</sup>, Lucía Granados<sup>2</sup>  
Facultad de Ciencias, Universidad Nacional Autónoma de México<sup>1</sup>  
Escuela Nacional de Antropología e Historia<sup>2</sup>



DH2018  
Ciudad de México | 26-29 JUNIO

## 1 Introducción

El libro **Méjico: del antiguo régimen a la revolución**, profundiza en la historia de la primera Revolución del siglo XX. La obra del historiador Francoise Xavier Guerra, se basó en una investigación prosopográfica y se convirtió en una referencia fundamental para el estudio del movimiento revolucionario en México. Su trabajo revela las relaciones y tensiones en la transición de la sociedad tradicional -un sistema heredado de la colonia- a un Estado moderno proveniente en gran medida de los ideales liberales de la revolución francesa. En el nuevo orden se erige al individuo como único repositorio de la razón y se sitúa como el eje central de la vida política, en detrimento de actores colectivos como la iglesia o los pueblos de indios. En ese sentido, rechaza cualquier vínculo que no surja de la libre asociación.

Sin embargo, la transición dista mucho de su modelo ideal, no solo por el arraigo de las instituciones antiguas, sino porque de manera tácita, incluso entre la élite liberal se conservan las lealtades familiares, y en los nuevos colectivos como las logias o las sociedades de oficios, frecuentemente se recrean las viejas dinámicas. Por ello consideramos que la elección de la estrategia prosopográfica como herramienta de aproximación al tema es sumamente fructífera ya que permite observar vínculos desestimados en el nuevo sistema, que no obstante, podrían contribuir a la comprensión de los movimientos en la cúpula política, que se caracterizó por un reducido número de personajes.

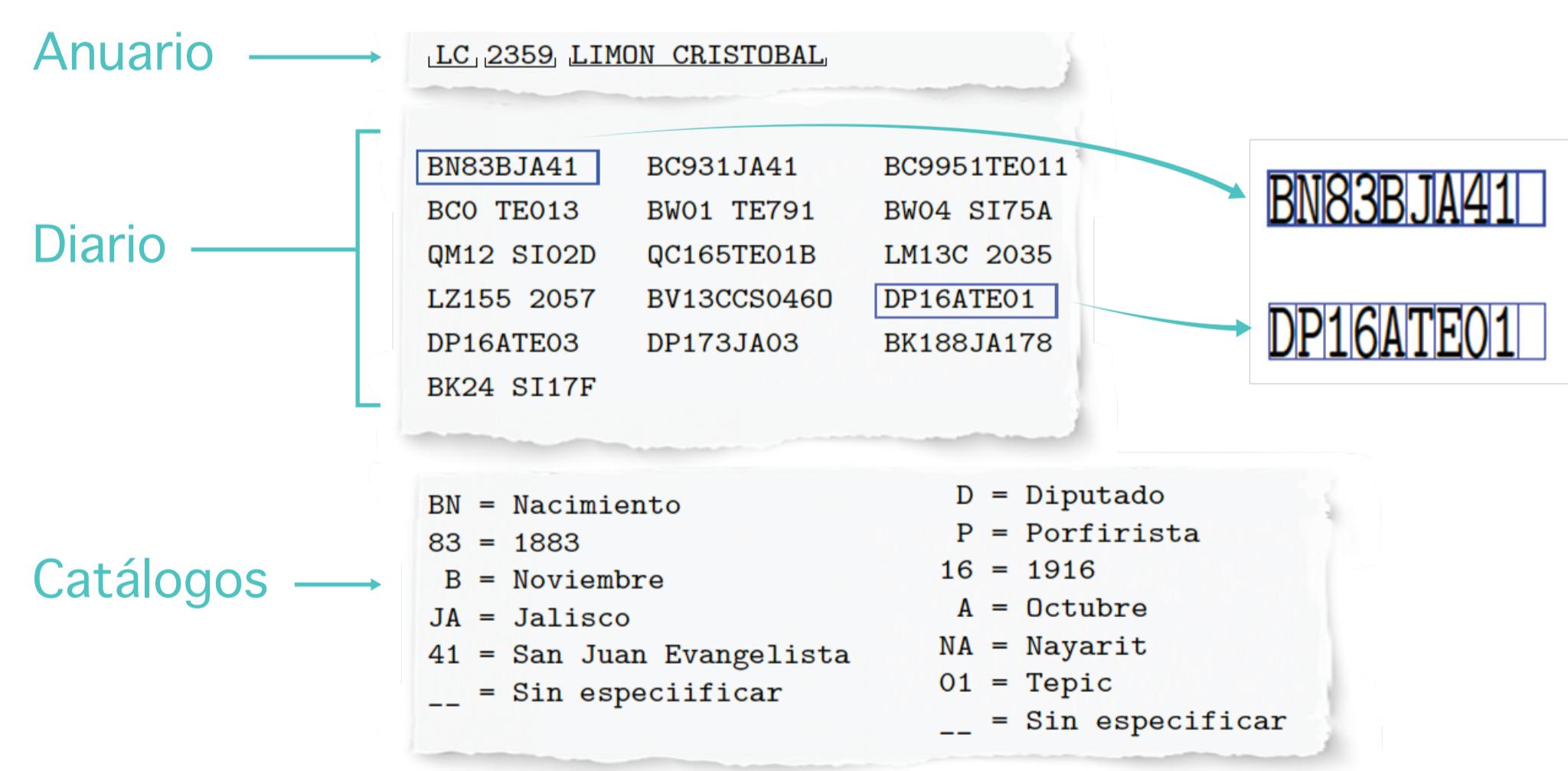
## 2 Objetivo

La base de datos de la obra fue almacenada originalmente en tres cintas magnéticas, y el único medio en que se encuentra disponible actualmente es impresa en los anexos del libro. El objetivo de este trabajo es la digitalización de la base para permitir la reproducción de los análisis del autor, así como la generación de nuevo conocimiento a partir del cruce de variables.

## 3 Organización de la base de datos

El cuerpo biográfico de la investigación está compuesto por más de tres mil actores sociales entre los que figuran individuos y colectividades, con aproximadamente cien mil datos asociados a los movimientos políticos. La información se sistematizó en tres secciones: el anuario, los catálogos y el diario.

- **El anuario** enlista la totalidad de los personajes, su registro consta de tres componentes, las iniciales, un número de identificación que es único para cada personaje y su nombre completo.
- **Los catálogos** sistematizan variables en cuarenta y dos categorías, las cuales pueden referir a ubicación -estado, ciudad o año-, a datos personales -nivel escolar, el grado de riqueza familiar o a la tendencia política- y datos relacionados con la vida política por -los cargos ocupados o el grado militar-.
- **En el diario**, para cada personaje, se conjunta la información de los catálogos en módulos, los cuales contextualizan sucesos como nacimiento, batallas o un cargo político ocupado, con fechas y lugares específicos. Los módulos constan de 10 caracteres, los primeros dos, son letras mayúsculas que identifican el suceso descrito, cada módulo es independiente lo cual permite enriquecer la base de datos con la captura de nuevos módulos para personajes ya establecidos.



## 4 Metodología

Se ocupó el framework OpenCV una biblioteca libre para reconocimiento visual artificial para trabajar sobre las hojas escaneadas de la obra, en este paso se procesaron las imágenes, y se detectó la estructura espacial del texto, para que posteriormente se procesaran pedazos de texto organizados.

En la digitalización se creó un programa en Python que ocupa Tesseract, y Pytesseract, ambas son bibliotecas de reconocimiento óptico de caracteres (OCR). El programa es un proceso semi-automatizado, que se detiene al detectar que la lectura OCR es inconsistente con las características de la base, entonces solicita a la persona que corrija la entrada manualmente para continuar con la lectura.

## 6 Resultado

Hasta el momento, se han completado el anuario y los catálogos. El diario está en proceso de digitalización, entre las dificultades para su avance encontramos que, en ocasiones, los valores de los módulos no corresponden con ninguna variable asignada en los catálogos, consideramos que pudieron haber sido resumidos en el proceso de publicación. Otro percance es la falta de información en los módulos que lo hacen carecer de sentido, por ejemplo el prefijo LX de los módulos debe referir a un vínculo, el siguiente espacio en el módulo designa qué tipo de vínculo y por último coloca el número de identificación del personaje al que esta relacionado, sin embargo, en múltiples ocasiones falta el dato del otro personaje. En ese sentido, consideramos que quizás sería de utilidad contar con la tesis doctoral del historiador Xavier Guerra, el documento que dio pie a la publicación del libro.

## Metodología

a) Hoja original

25	26
ME 2095 INFANTEZON, JESÚS LINO P.	BN 2218 BOLÍVAR, V. GREGORIO
CJ 2095 MARQUES, JOSÉ MANUEL	BP 2219 BONILLA, V. CLAUDIO A. J.
GF 2102 GUNZBURG, FABRIZIO	CN 2220 BONILLA, V. JOSÉ
GE 2102 GARCIA AZO, ARTURO	CR 2221 CANO, NICOLAS
GE 2102 GARCIA AZO, JUAN	CS 2222 CASTAÑEDA, JUAN
MA 2111 MARTINEZ, ALFREDO JOSE	CF 2223 CASTAÑEDA Y CASTAÑEDA, RAMON
SM 2114 SANTOS ALONSO, MANUEL M.	CM 2224 CASTILLO, JOSÉ
SM 2114 SANTOS ALONSO, MANUEL M.	CM 2245 CEDEÑO, MARCELINO
SR 2122 SISTERS, ANTONIO INNOCENTIO	CO 2246 CEDEÑO, MARCELINO
SR 2122 SISTERS, ANTONIO INNOCENTIO	CR 2247 CEDEÑO, MARCELINO
TM 2127 TRIANA GUERRA, MARTIN	CS 2248 CELESTE DE LA CRUZ, RAFAEL
VS 2128 VALLADARES, JAVIER	CF 2249 CERDÁ, JOSÉ
LA 2128 VILLALBA RODRIGUEZ, ANTONIO I.	CM 2250 CERDÁ, JOSÉ
LA 2128 VILLALBA RODRIGUEZ, ANTONIO I.	CO 2251 CERDÁ, JOSÉ
PA 2143 PORTAS, ANTONIO	CR 2252 CERDÁ, JOSÉ
AL 2146 AGUILAR LAMBERT, FEDERICO	CS 2253 CERDÁ, JOSÉ
AL 2146 AGUILAR LAMBERT, FEDERICO	CF 2254 CERDÁ, JOSÉ
MJ 2146 MUÑEZ, JAVIER, FEDERICO	CM 2255 CERDÁ, JOSÉ
CF 2150 COSS, FRANCISCO	CO 2256 CERDÁ, JOSÉ
AM 2152 ACOSTA, MIGUEL ANGEL	CR 2257 CERDÁ, JOSÉ
AM 2152 ACOSTA, MIGUEL ANGEL	CS 2258 CERDÁ, JOSÉ
DI 2154 LIMA, CRISTOBAL	CF 2259 CERDÁ, JOSÉ

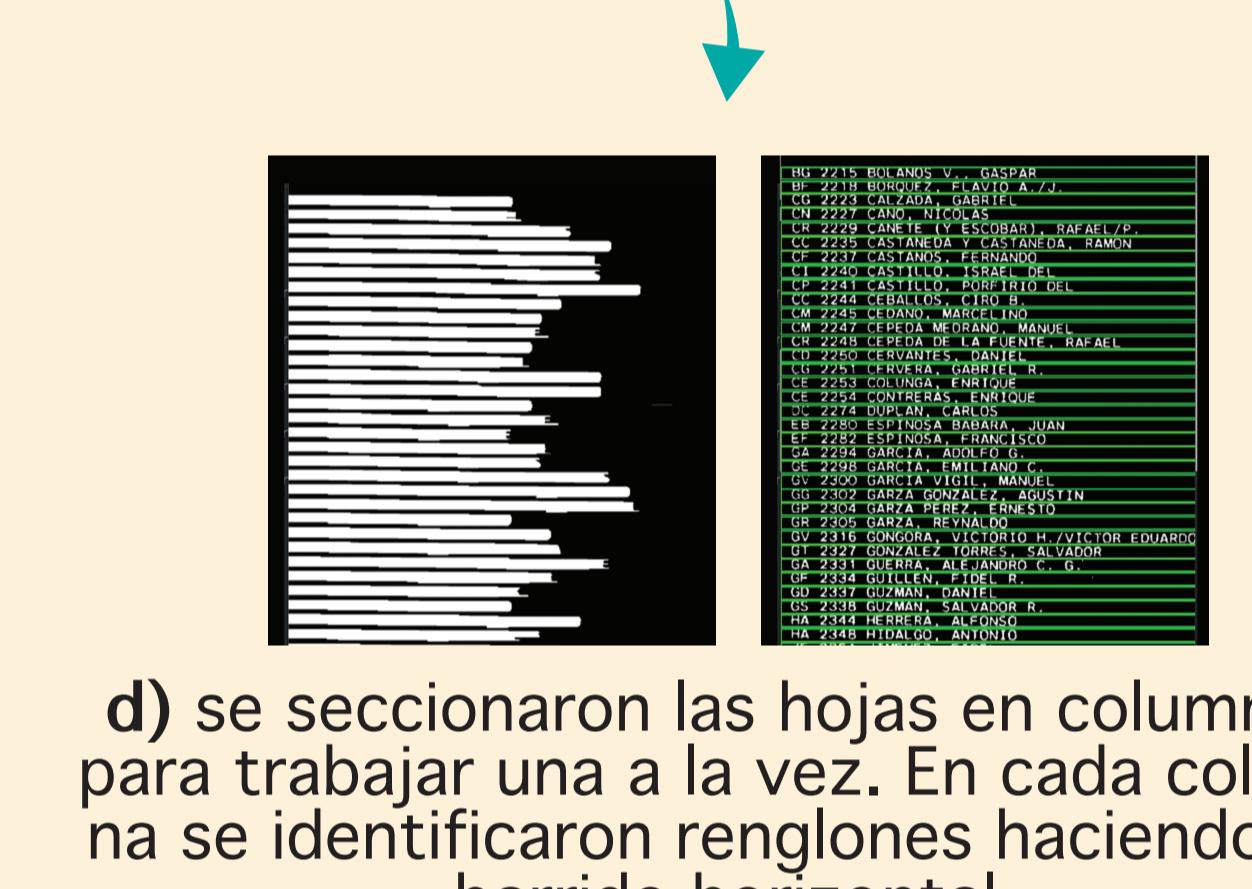
b) se mejoró la calidad de las imágenes, haciéndola binaria

25	26
ME 2095 INFANTEZON, JESÚS LINO P.	BN 2218 BOLÍVAR, V. GREGORIO
CJ 2095 MARQUES, JOSÉ MANUEL	BP 2219 BONILLA, V. CLAUDIO A. J.
GF 2102 GUNZBURG, FABRIZIO	CN 2220 BONILLA, V. JOSÉ
GE 2102 GARCIA AZO, ARTURO	CR 2221 CANO, NICOLAS
GE 2102 GARCIA AZO, JUAN	CS 2222 CASTAÑEDA, JUAN
MA 2111 MARTINEZ, ALFREDO JOSE	CF 2223 CASTAÑEDA Y CASTAÑEDA, RAMON
SM 2114 SANTOS ALONSO, MANUEL M.	CM 2224 CASTILLO, JOSÉ
SM 2114 SANTOS ALONSO, MANUEL M.	CM 2245 CEDEÑO, MARCELINO
SR 2122 SISTERS, ANTONIO INNOCENTIO	CO 2246 CEDEÑO, MARCELINO
SR 2122 SISTERS, ANTONIO INNOCENTIO	CR 2247 CEDEÑO, MARCELINO
TM 2127 TRIANA GUERRA, MARTIN	CS 2248 CELESTE DE LA CRUZ, RAFAEL
VS 2128 VALLADARES, JAVIER	CF 2249 CERDÁ, JOSÉ
LA 2128 VILLALBA RODRIGUEZ, ANTONIO I.	CM 2250 CERDÁ, JOSÉ
PA 2143 PORTAS, ANTONIO	CO 2251 CERDÁ, JOSÉ
AL 2146 AGUILAR LAMBERT, FEDERICO	CR 2252 CERDÁ, JOSÉ
MJ 2146 MUÑEZ, JAVIER, FEDERICO	CS 2253 CERDÁ, JOSÉ
CF 2150 COSS, FRANCISCO	CF 2254 CERDÁ, JOSÉ
AM 2152 ACOSTA, MIGUEL ANGEL	CM 2255 CERDÁ, JOSÉ
AM 2152 ACOSTA, MIGUEL ANGEL	CO 2256 CERDÁ, JOSÉ
DI 2154 LIMA, CRISTOBAL	CR 2257 CERDÁ, JOSÉ

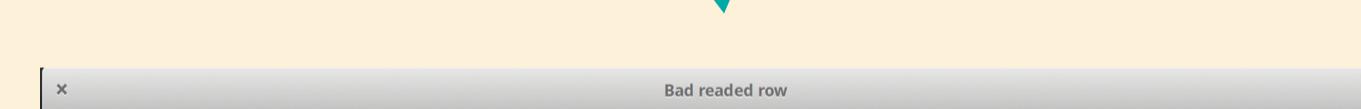
### Proceso para el Anuario



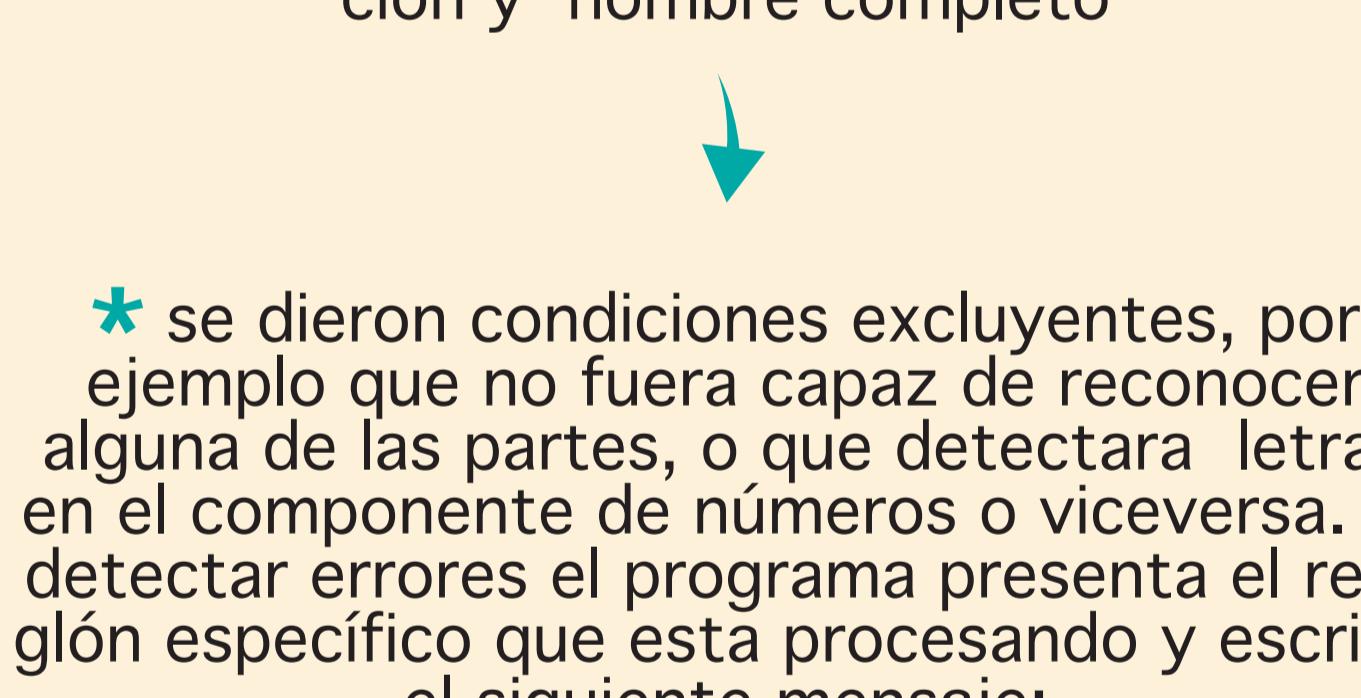
c) se identificaron las columnas en el texto mientras el barrido de las letras hacia abajo, lo cual permitió identificar los tres componentes



d) se seleccionaron las hojas en columnas para trabajar una a la vez. En cada columna se identificaron renglones haciendo un barrido horizontal

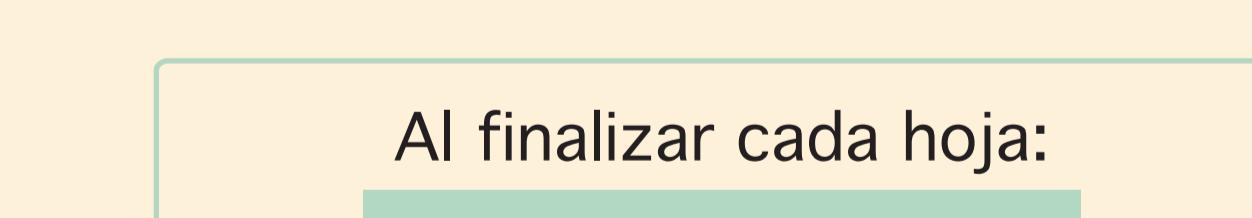


e) Se ejecuta el OCR con tesseract y en cada renglón de las columnas se identificaron los tres componentes del anuario: iniciales, numero de identificación y nombre completo



\* se dieron condiciones excluyentes, por ejemplo que no fuera capaz de reconocer alguna de las partes, o que detectara letras en el componente de números o viceversa. Al detectar errores el programa presenta el renglón específico que está procesando y escribe el siguiente mensaje:

'InPlease, help me to fix the following errors (• ω •):'



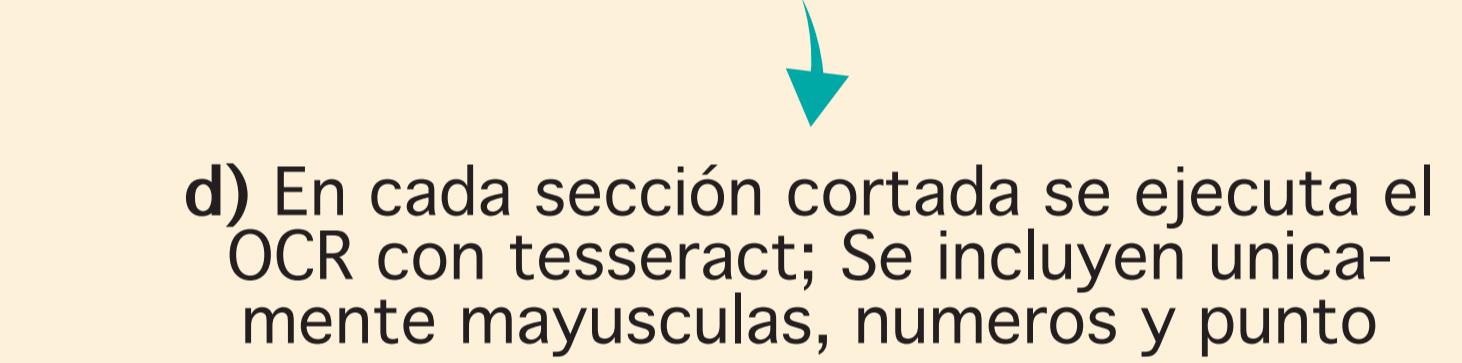
El anuario el diario y los catálogos se almacenaron en archivos .csv

Debido a la estructura modular del diario, los renglones, columnas y espacios en blanco son significativos.

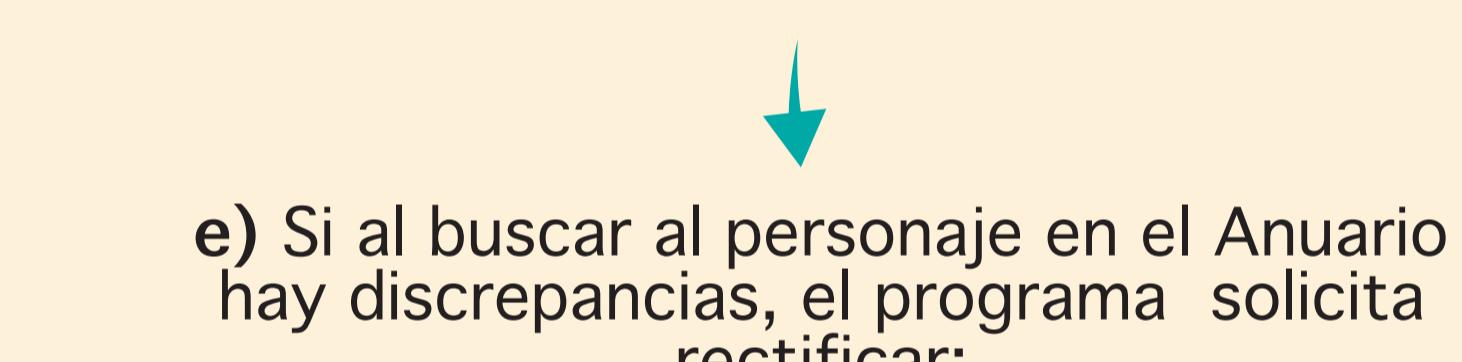


c1) se identificaron renglones haciendo un barrido horizontal.

c2) aprovechando la indentación de los módulos respecto a los personajes, se distinguieron los personajes de sus módulos

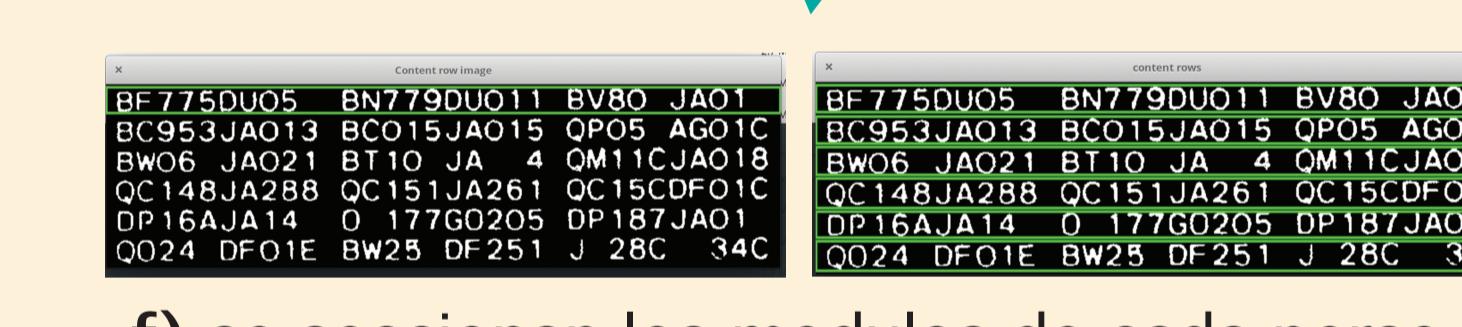


d) En cada sección cortada se ejecuta el OCR con tesseract; Se incluyen únicamente mayúsculas, números y punto

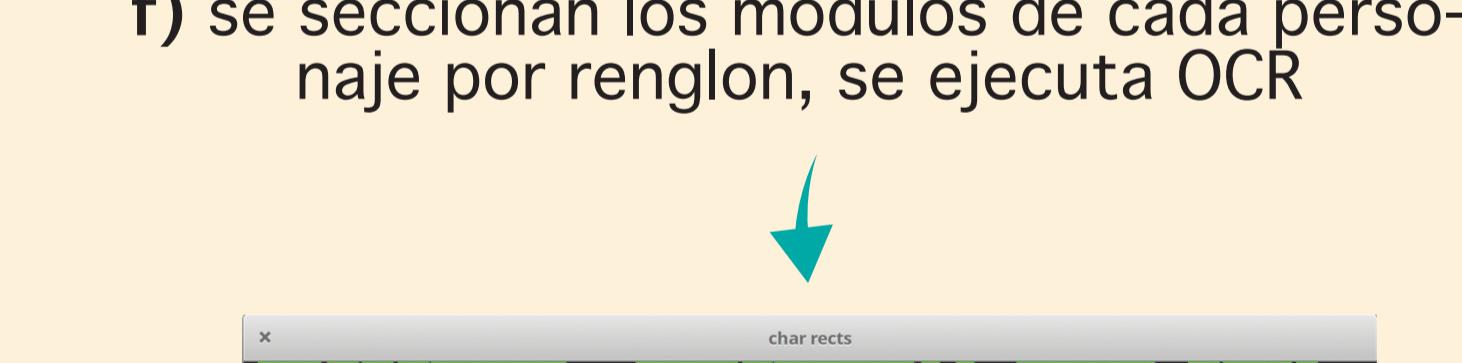


e) Si al buscar al personaje en el Anuario hay discrepancias, el programa solicita rectificar:

'Help me to choose one: '

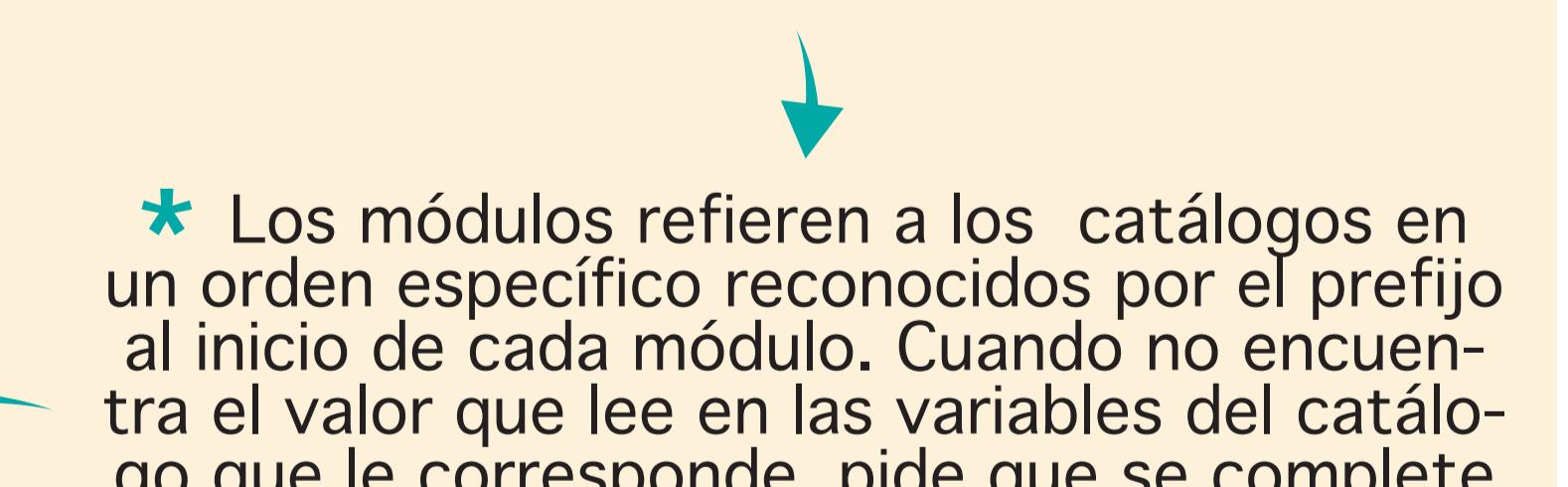


f) se seleccionan los módulos de cada personaje por renglón, se ejecuta OCR



g) cada renglón contiene 3 módulos con 10 caracteres, separados por un espacio (máximo 33 caracteres) y los espacios en blanco se leen como '\_'. Cuando encuentra una excepción señala el error con el siguiente mensaje

'Row only should have 3 or less modules.. Help me to fix it: '



\* Los módulos refieren a los catálogos en un orden específico reconocidos por el prefijo al inicio de cada módulo. Cuando no encuentra el valor que lee en las variables del catálogo que le corresponde, pide que se complete con el siguiente mensaje

'Not found value: CD on catalog: 3 in: QCI15CD-FO1. Help me to fix it: '

## 7 Perspectivas

Se planea poner a disposición la base de datos en un sitio web, para que los usuarios puedan realizar queries inéditos e incluso alimentar la base completando o creando módulos. Consideramos que este trabajo ofrecerá la oportunidad de profundizar en el conocimiento sobre los períodos del Porfiriato y la Revolución Mexicana a partir de información sistematizada por el historiador Xavier Guerra.

### Contacto

diegomontesinos@ciencias.unam.mx

luciagranadosriveros@gmail.com