

Equipo de digitalización: Proyecto Incipit

Gabriela Striker

Agustina Miguens

Ileana Campagno

Objetivos

La labor de digitalización de las revistas *Incipit* publicadas por el Seminario de Edición y Crítica Textual (SECRIT/CONICET) tiene por objetivo la preservación de este material científico y su difusión en una plataforma web para su consulta en línea.

Duración

El proceso se desarrolló desde mediados de abril a mediados de diciembre de 2017.

Resumen del trabajo

El trabajo constó de las siguientes tareas: la digitalización propiamente dicha de las revistas *Incipit* publicadas entre 1981 y 2009 y su pos-proceso para la confección de los pdfs finales; la revisión general de estos; la recaptura de las páginas que presentaban problemas de focalización y de texto y/o foto mutilados; la aplicación del reconocimiento óptico de caracteres; la aplicación del programa Scantailor para la optimización de la calidad de las imágenes y la ejecución necesaria de un nuevo pos-proceso; la compaginación final de las recapturas y la eliminación de las páginas en blanco del total del material escaneado mediante el programa Smallpdf; la revisión final y la presentación del producto digitalizado acompañado de este breve informe.

Descripción del proceso

Luego de recibir la capacitación por parte de Matías y Juan Pablo (una reunión de presentación y tres más de explicación del proceso), y de escanear entre las tres una *Incipit* de práctica, empezamos yendo de a dos a digitalizar, para poder solventar posibles dudas o problemas que nos surgieran durante la digitalización. Más adelante, cuando ya tuvimos más internalizado el proceso, fuimos en forma individual.

Nos armamos un Excel en Google Drive en el cual fuimos completando una planilla de revisión cada vez que terminábamos de digitalizar. Siempre había dos revisoras e íbamos anotando allí mismo los comentarios.

La mayor dificultad con la que nos topamos en la fase de escaneado fue la visualización de las capturas; no se les puede aplicar zoom y por esto no se ve con claridad si estas salen enfocadas y/o si se cortan los márgenes. Se ralentiza el trabajo ya que hay que entrar al programa donde están todas las capturas y detenerse en cada una para definir estas cuestiones. Del problema anterior derivó, en gran medida, la recaptura de numerosas páginas ya que durante la primera revisión general registramos páginas con texto desenfocado/borroso, márgenes cortados, imágenes que no se veían bien (por lo general salían muy oscuras) y rayas en los márgenes, además de páginas salteadas y repetidas. Otro problema con las imágenes fue un caso en el que directamente se veían completamente deformadas debido al ángulo de captura. Esto se solucionó modificando el mismo. En el caso de las páginas que nos habíamos salteado y aquellas que habían salido con el texto desenfocado o borroso, o con el margen cortado, las volvimos a escanear y a posprocesar.

La etapa del posproceso de estas recapturas nos demandó bastante tiempo porque debíamos posprocesar nuevamente todo el volumen y a veces solo por unas pocas páginas por cada ejemplar. Cabe agregar a esto que el pdf nuevo que se genera no reemplaza automáticamente al anterior, por lo cual debíamos atender a eliminarlo previamente.

Luego utilizamos mucho un programa online que se llama Smallpdf para volver a compaginar las revistas. Con este programa insertamos las páginas que faltaban, quitamos las que no se veían bien y las reemplazamos por las adecuadas, y también eliminamos todas las páginas en blanco, que nos habían solicitado quitar para que los pdfs finales quedaran más prolijos. Todo este trabajo con el Smallpdf lo realizamos desde nuestras casas una vez que ya teníamos los pdfs de las *Incipit*.

Las imágenes que se veían mal tuvimos que arreglarlas utilizando otro programa que nos enseñó a usar Juan Pablo en dos tardes de capacitación más, ya avanzado el proceso de digitalización, que se llama Scantailor (y que forma parte del “paquete de digitalización” que los chicos programaron, por lo cual no se podía hacer desde casa, sino que teníamos que ir sí o sí a la oficina). Las *Incipit* las digitalizábamos en blanco y negro, por lo cual muchas imágenes quedaban empastadas. Las que se veían bien las dejamos así, pero las que se veían mal las pasamos a “mixed” con el Scantailor (es decir, a escala de grises), con lo cual mejoraba la calidad notablemente. Simultáneamente tuvimos en cuenta cómo se hallan en los originales con el fin de preservar lo mejor posible el criterio editorial. Asimismo, en ciertos casos, cuando incluso los textos o notas al pie no se veían bien por estar muy clara la impresión en el original, decidimos aplicar Scantailor (cambiar de “blanco y negro” a “mixed”) para rescatar las palabras puesto que, de otro modo, no es posible editarlas con este programa. Luego había que volver a realizar el posproceso e insertar las páginas utilizando Smallpdf. En cuanto a las rayas que salían en algunos márgenes, solo quitamos algunas que interferían en los textos, las demás decidimos dejarlas porque no afectan para nada la lectura de los artículos.

Con respecto al programa Scantailor, notamos que los comandos no nos resultaban tan intuitivos como los del Dalclick, donde además contábamos con un menú con los atajos. Además, cada vez que queríamos guardar una o varias imágenes a las que les habíamos aplicado Scantailor, debíamos cerrar el programa. Esto resulta tedioso cuando un volumen presenta gran cantidad de imágenes pero, para no correr el riesgo de perder los cambios, tenemos que hacerlo. Por ejemplo, cada 5 cambios realizados, debíamos cerrar Scantailor y abrirlo nuevamente para continuar con el mismo volumen. Aún así, logramos corregir todas las fotos, cuadros, retratos, reproducciones facsimilares de manuscritos y demás imágenes empastadas.

A partir de las observaciones sobre las fases de digitalización, posproceso y recaptura sugerimos:

- ampliar la pantalla de visualización de las capturas al momento de digitalizar, que se pueda aplicar zoom.
- programar el posproceso de unas pocas páginas para evitar posprocesar todo el volumen de nuevo.
- que el pdf nuevo que se genera al posprocesar un volumen reemplace automáticamente al anterior sin necesidad de eliminarlo previamente.

A partir de las observaciones sobre la fase de aplicación de Scantailor sugerimos:

- generar un instructivo/manual con las funciones básicas.
- habilitar un icono “guardar” como utilizamos, por ejemplo, en Word así no corremos riesgo de perder ningún cambio aplicado a las imágenes. Además, resulta más práctico y dinamiza el trabajo.
- evaluar la posibilidad de la edición de palabras con este programa.

Cabe agregar que las etapas de posproceso y Scantailor podrían desarrollarse desde nuestras computadoras personales sin necesidad de trasladarse al lugar donde se encuentre el escáner y la pc central.

Luego de obtener los pdfs finales realizamos el reconocimiento óptico de caracteres (OCR). Al margen de las erratas que observamos después de su aplicación en los títulos de las obras que se encuentran en cursiva o subrayados, en los nombres de los

autores de los artículos y en las referencias bibliográficas, el OCR funcionó muy bien para los textos. Aclaramos que no nos abocamos de manera más profunda a la corrección de estas erratas porque, por un lado, no disponemos de un editor para solucionar estas contingencias y, por otro lado, esta tarea nos hubiera demandado mucho más tiempo.

Con respecto a la fase de revisión final, procedimos a la división de la tarea de la siguiente manera:

- Incipit* I a IX a cargo de Ileana y Agustina;
- Incipit* X a XVIII a cargo de Gabriela e Ileana;
- Incipit* XIX a XXIX a cargo de Agustina y Gabriela.

Estos son los resultados de la revisión por pares de acuerdo con algunos criterios básicos que establecimos para esta fase (1. Focalización y legibilidad; 2. Hojas en blanco; 3. Paginación; 4. Márgenes y rayas; 5. Imágenes):

1. Focalización y legibilidad:

1.1. No encontramos problemas de focalización.

1.2. Legibilidad:

Incipit I: imagen torcida (*Crónicas de los Reyes de Castilla*) pero legible entre las pp. 70 y 71.

Incipit IV: la p. 48 está un poco borrosa (no desenfocada), pero es legible.

Incipit V: el renglón superior de la p. 183 es casi ilegible.

Incipit VIII: la p. 18 está un poco borrosa (no desenfocada) al final; a la p. 92 se le aplicó Scantailor para que se pudieran visualizar las líneas superiores.

Incipit XXVII: se le aplicó Scantailor a la nota al pie de la p. 288 para que se pudiera leer.

2. Hojas en blanco: no encontramos hojas en blanco para eliminar (todas fueron eliminadas durante las fases anteriores).

3. Paginación:

3.1. No hallamos páginas repetidas (ya fueron eliminadas durante las fases anteriores).

3.2. No observamos desorden entre las páginas, lo que podría haberse producido como consecuencia de las recapturas y su posterior compaginación.

3.3. *Incipit* X: si bien la numeración está en orden, registramos que el texto de Jorge Ferro es incoherente entre las páginas 65 y 66. Podría/n faltar una o más páginas entre las mencionadas. El error se encuentra en el original empleado.

4. Márgenes y rayas:

4.1 Observamos líneas en los márgenes, producto del sistema de digitalización, que no representan obstáculo alguno para la lectura, ya que no cortan ni rozan los textos.

4.2. Encontramos márgenes cortados en las siguientes revistas:

Incipit XI: p. 121 y p. 133; no impiden la lectura pero la dificultan, especialmente para el caso de la p. 133.

Incipit XV: pp. 23, 25, 27, 83, 85 y 87; no impiden la lectura pero la dificultan bastante.

Incipit XXV-XXVI: p. 396 y p. 398 (en este caso, se advierte en el encabezado que acompaña a la foto superior o "lámina 4")

5. Imágenes: no hallamos imágenes empastadas (todas las que presentaban esta característica fueron mejoradas mediante la aplicación de Scantailor en la etapa anterior).

Conclusiones

Más allá de las dificultades encontradas –asociadas a nuestra primera experiencia con este tipo de escáner en forma de libro (abertura en V) y con los programas que acompañaron el proceso así como al estado de los originales, en particular, de los volúmenes mecanografiados–, consideramos que el producto final, que consta de 27 revistas digitalizadas, es de alta calidad. De todas maneras, dejamos a criterio de los directores del proyecto los cambios que creyeren conveniente realizar al trabajo logrado hasta la fecha.

Apéndice (ficha entregada a Matías y Juan Pablo)

Observaciones y sugerencias

Fase 1: Digitalización y posproceso

-las cámaras se tildaban frecuentemente

-ampliar pantalla de visualización de las capturas al momento de digitalizar, que se pueda aplicar zoom ahí mismo (no se ve si las capturas salen enfocadas o no, si los márgenes salen cortados o no, y se hace necesario invertir tiempo en entrar al programa). Incluso haciendo esto, en ocasiones quedaron márgenes o imágenes cortados y fue necesario recapturar volúmenes enteros.

-que sea posible programar el posproceso de unas pocas páginas para evitar posprocesar todo el volumen de nuevo.

-cuando se vuelve a posprocesar un volumen, el pdf nuevo que se genera podría reemplazar automáticamente al anterior sin necesidad de eliminarlo previamente. ¿Habría alguna razón por la cual no fuera conveniente que se ejecutara de esta manera?

-elementos necesarios para trabajar en una futura etapa de digitalización:

-herramienta para ajustar y aflojar las mariposas de los brazos que sostienen la luminaria del escáner, manualmente se dificulta bastante (pasó que los brazos se caen y entorpecen la dinámica del trabajo).

- “maderita”, en términos de Juan Pablo, o algún otro elemento para colocar en la cuna y que levante el libro cuando este es muy voluminoso.

- Posibilidad de variar el ángulo de apertura de la cuna.

Fase 2: Scantailor

-Generar un instructivo/manual con las funciones básicas: los comandos no eran tan intuitivos como los del Dalclick, donde además contábamos con un menú con los atajos.

-¿se puede editar palabras, frases, notas al pie y demás con este programa? ¿Cómo?

-cerrar el programa cada vez que uno quiere guardar una o varias imágenes a las que se les ha aplicado Scantailor resulta riesgoso (se puede perder información fácilmente en caso de que se tildé el programa). Sería muy práctico hacer click directamente en un icono "guardar" como utilizamos, por ejemplo, en Word así no corremos riesgo de perder ningún cambio.

-sería genial aplicar Scantailor desde una pc personal en nuestras casas.