

Guía para la digitalización accesible de materiales bibliográficos

Autor: Daniel Jesús García - ex delegado estudiantil de la Comisión de Inclusión Educativa

Índice

I. Introducción.....	1
II. Cuestiones legales de la digitalización de bibliografía con fines de accesibilidad.....	2
III. Lectores de pantalla y software OCR.....	2
IV. 3 pasos para digitalizar un libro de manera accesible.....	3
V. Convirtiendo PDFs en formatos accesibles.....	4
Anexo I. Algunos programas OCR	5

I. Introducción

Hace un siglo atrás, era impensada la posibilidad de que una persona con alguna dificultad para leer en papel (discapacidad visual, motriz, etc.), accediera al texto sin ayuda de otra persona.

Con la aparición de las tecnologías de la información y comunicación (TICS), esta realidad comenzó a cambiar. Con la aparición de las computadoras, se empezó a plantear el problema de cómo lograr que las personas ciegas o con otra dificultad pudieran utilizar estos dispositivos.

Para dar una respuesta a este problema, se desarrollaron aplicaciones llamadas “lectores de pantallas”. Las mismas permitían y permiten que una persona ciega o con baja visión utilice un ordenador, y más actualmente, dispositivos móviles.

Con el desarrollo de Internet el libro comenzó a circular en formatos digitales y ya no únicamente en papel. Estos acontecimientos plantearon algo inesperado: la posibilidad de acceder a los textos para las personas que no podían hacerlo en el clásico formato de libro.

Hoy en día sabemos que muchos libros están disponibles para descargar en diversas webs de manera paga o gratuita.

Sin embargo, existe bibliografía que no se encuentra en formato digital, lo cual supone un problema para el acceso a la misma por parte de personas que únicamente pueden leer o escuchar de esta manera.

Es de este modo que llegamos a la digitalización de libros para tornarlos accesibles. El propósito de este documento es proporcionar una serie de conceptos básicos y herramientas informáticas a la hora de tornar accesible diversos materiales de estudio.

II. Cuestiones legales de la digitalización de bibliografía con fines de accesibilidad

En el año 2013, la “organización mundial de la propiedad intelectual” (OMPI), impulsó el: “Tratado de Marrakech para facilitar el acceso a las obras publicadas a las personas ciegas, con discapacidad visual o con otras dificultades para acceder al texto impreso”. En dicho tratado, se propone que se creen excepciones al derecho de autor con fines de accesibilidad.

Uruguay ratificó este tratado en el año 2014, mediante la Ley 19.262.

Este tratado y su ratificación por nuestro país otorgan un marco legal a la digitalización de contenido bibliográfico, siempre que esta sea con fines de accesibilidad.

III. Lectores de pantalla y software OCR

Pasemos ahora a considerar algunas aplicaciones informáticas en materia de accesibilidad.

Lector de pantalla

Un lector de pantalla es un programa que permite a un usuario ciego o con baja visión manejar la computadora utilizando únicamente el teclado. El lector de pantalla lee en voz alta el contenido que se esté mostrando en el monitor.

De esta manera, el usuario puede manejarse a nivel de sistema operativo (manejar archivos, carpetas, etc.), de procesador de texto (leer y escribir), de internet (manejar correo electrónico, redes sociales, y todo tipo de páginas webs), y manejar la computadora a nivel general.

En sistemas operativos Linux, el lector de pantalla más utilizado es “Orca”. En sistemas operativos Windows, hay dos opciones principales: una de ellas es el “NVDA”, un lector de pantalla gratuito que se ha venido perfeccionando con el correr de los años. El otro es el conocido “Jaws”, un lector de pantalla de carácter privado y muy potente en cuanto a funcionalidades. En sistemas operativos Windows, los lectores de pantalla no suelen leer los documentos pdf, por lo que se deben convertir a otros formatos como Word, rtf o texto simple (txt). Veremos como hacer estas conversiones de pdf en un apartado posterior

OCR

De las siglas en inglés Optical Character Recognition (reconocimiento óptico de caracteres), un OCR es un software que nos permite reconocer el contenido textual presente en imágenes, sean en archivos de este tipo o en documentos pdf. El OCR realiza una lectura de las imágenes en búsqueda de texto para extraerlo y a continuación, generar un documento que se pueda leer y editar en un procesador de texto.

Existen aplicaciones gratuitas y pagas para realizar este reconocimiento óptico de caracteres. También existen páginas en internet que realizan la tarea subiendo el documento pdf a la misma y descargándolo luego en formato de texto. En el anexo I se incluye una lista de programas OCR más utilizados, tanto gratuitos como pagos.

IV.3 pasos para digitalizar un libro de manera accesible

Paso I: escaneo

Para que el libro pueda ser leído en forma digital por un lector de pantalla, es un requisito indispensable que esté bien escaneado. Para lograr este cometido, el libro debe estar en buenas condiciones materiales. Por ejemplo, un libro subrayado, manchado o en malas condiciones será de difícil acceso para un OCR, y por lo tanto, los lectores de pantalla leerán palabras cortadas o sin sentido.

Este paso finaliza con la producción por parte del escáner de un documento pdf con las imágenes correspondientes a cada página del libro, o con un conjunto de imágenes en una carpeta, donde cada imagen es una página del mismo.

El texto que contienen estas imágenes debe estar claro y ser legible sin problemas. Esto nos da una pista de si el proceso de escaneo marchó bien o si se produjo alguna falla, sea en el estado del libro o en el escaneado mismo.

Paso II: reconocimiento de texto OCR

Una vez que tenemos las páginas del libro en imágenes, sea en pdf (que es lo más práctico), o en imágenes sueltas, el paso que sigue es convertir estas imágenes a texto para que un lector de pantalla pueda acceder al mismo.

Para esto, utilizaremos un OCR. Podemos instalar uno en nuestra computadora, utilizar alguno disponible en la web, o si nuestro escáner trae la opción de OCR, escanear ya con dicha opción. En cualquier caso, la idea siempre será convertir el documento pdf o las imágenes a un documento de texto, sea este en formato Microsoft Word o en algún otro formato de texto, como ser txt o rtf en sistemas operativos Windows.

Paso III: verificación

Una vez que ya tenemos un documento de texto, es recomendable verificar como ha ido el proceso y si se ha perdido o modificado texto del libro original. Para lograr este objetivo, se puede leer el documento de texto a la par que el libro material y verificar si las palabras han quedado bien digitalizadas. Generalmente basta con leer las primeras páginas y comprobar si las palabras del documento de texto existen en nuestro idioma y si el texto es coherente. Si el proceso de escaneo y el OCR han marchado bien, el documento de texto producido debería ser bastante preciso.

Esto en lo referido a digitalizar libros de texto. Cuando en dichos libros hay imágenes, esquemas o algún componente gráfico, el OCR suele no funcionar con dichos elementos o funcionar muy mal. En este caso, se debe evaluar la posibilidad de explicar las imágenes y los esquemas en formato de texto. Es decir, poner en palabras las imágenes cuando esto es posible. En imágenes complejas se deben buscar otras estrategias junto a la persona para hacer accesible tales imágenes. Por ejemplo, la creación de “imágenes en tres dimensiones” puede facilitar que la persona comprenda utilizando el tacto, por ejemplo, los componentes de una célula neuronal. Para esto se pueden utilizar diversos materiales (plasticina, porcelana fría, cascolas de colores, etc). Una solución más sofisticada y actual es el uso de una impresora 3d.

Desde luego, esto debe ir acompañado de una explicación sobre que significa cada forma. Luego la persona podrá unir ese nombre con esa forma e ir y venir entre esta y el texto.

V. Convirtiendo PDFs en formatos accesibles

Como se ha mencionado antes, los lectores de pantallas no suelen leer en formato pdf. Si el documento pdf contiene imágenes con texto encapsulado, simplemente podemos empezar en el paso II del apartado anterior, es decir, aplicando un OCR y verificando luego el proceso.

Es frecuente el uso de documentos pdf que ya contienen texto seleccionable. Por ejemplo, los artículos académicos publicados en una revista científica suelen estar en formato pdf, pero el texto se puede seleccionar con el ratón de la computadora. En estos casos, lo único que se necesita es seleccionar el documento, copiar el texto y pegarlo en un procesador de texto. También se puede utilizar la aplicación gratuita “Balabolka”, que posee la opción de convertir el documento pdf a formato txt, además de poder convertir también el pdf en formato de audio para ser reproducido en cualquier dispositivo. Para utilizar Balabolka con un archivo pdf, simplemente se hace click con el botón derecho en el archivo, se va al submenú Balabolka y se accede allí a estas funciones. Estas tecnologías que convierten el texto en audio se denominan TTS (Text to Speech) y constituyen otra buena herramienta a tener en cuenta a la hora de hacer accesible un texto.

Anexo I: Algunos programas OCR

Existen programas gratuitos y pagos para convertir imágenes con texto a documentos editables. Algunas de las opciones gratuitas son:

FreeOCR, SimpleOCR, GT Text, Me OCR Converter, Image to OCR Converter, etc.

Dentro de las opciones pagas, uno de los más conocidos y utilizados en español es el ABBYY Fine Reader. Este software es altamente recomendable. Otras opciones son:

OmniPage Standard, Readiris, Soda PDF, entre otros.