

Tutorial para escanear y compartir

Versión 1.07

**Escrito por Vadimir T. T. - Traducido al inglés por A.
Traducido al español por Armando.**

2008

Índice

1	Introducción	2
2	Escanear un libro	4
2.1	Configuración del IrfanView para escaneo	6
2.2	Trabajo manual al momento de escanear	8
3	Procesar escaneados con el ScanKromsator	10
3.1	Ejecución en borrador	11
3.2	Configuración de opciones	13
3.3	Ejecución definitiva	16
4	Procesar figuras en color y fotografías	17
5	Codificar escaneados en DJVU	18
6	Crear capas de texto con OCR	21
7	Agregar cubiertas y láminas de color al libro	23
8	Agregar hiperenlaces y marcas	24
A	¿Dónde descargar los programas?	27

Nota del traductor: Este documento fue escrito originalmente en ruso. El traductor agregó algunas instantáneas de IrfanView en inglés y otros detalles menores. Las instantáneas para del programa Djvu Hyperlinks Editor (Editor de hiperenlaces para Djvu), quedaron en ruso porque el programa no ofrece otra opción.

1. Introducción

Este es un mini tutorial acerca del escaneo de libros y la creación de sus archivos en alta calidad. El tutorial es para todo principiante que quiera hacer libros electrónicos de buena calidad y no sabe por dónde empezar. Existen muchas maneras de obtener buenos resultados mediante el escaneo; este documento le enseña un camino bastante fácil. En el tutorial se muestran imágenes de apoyo para cada paso y supone que el usuario está familiarizado con Windows. Seguramente necesitará descargar e instalar algunos programas (vea dónde encontrarlos en el apéndice A).

Nos enfocaremos principalmente a la digitalización de viejos libros de ciencias, matemáticas o técnicos. Para este tipo de libros, no es recomendable el sistema de OCR (Reconocimiento Óptico de Caracteres), porque dichos libros contienen abundantes ecuaciones, diagramas, gráficas, etc. La mejor solución es escanearlo y crear imágenes de todas sus páginas. La mayoría de estos libros casi siempre están impresos en blanco y negro, quizá con unas cuantas páginas ilustradas en color. Con libros de estas características, se logra la mayor calidad de escaneo si se generan todas o casi todas las imágenes a **600dpi** en blanco y negro.¹ Esto quiere decir que deberá escanear directamente en blanco y negro a 600dpi o en **escala de grises a 300dpi** y luego procesar lo escaneado para convertirlo a blanco y negro de 600dpi.² Si los libros tienen algunas páginas con ilustraciones en color, habrá que escanearlas por separado, a 300dpi en el modo color de 24 bits. Esto mismo se aplica para las cubiertas en color que quiera incluir.

Considere esto:

- **¡Nunca** escanee a 300dpi en blanco y negro! La calidad obtenida jamás será tan buena como la obtenida al escanear en escala de grises a 300dpi. Esta recomendación es válida tanto si sigue los pasos de este tutorial o los de cualquier método equivalente.
- En la mayoría de los escáneres, escanear en **escala de grises a 300dpi** es *¡exactamente igual de rápido!* que escanear a 300dpi en blanco y negro o en cualquier otra baja resolución. No ahorrará tiempo si escanea en blanco y negro a 300dpi o 200dpi, en vez de hacerlo en escala de grises a 300dpi, pero sí perderá *mucha* calidad.
- Escanear en **escala de grises a 300dpi** produce archivos intermedios de gran tamaño, los cuales después de procesados generarán archivos finales DJVU muy pequeños. En cambio,

¹ Si no sabe qué significa 600dpi: esta es la llamada *resolución* de la imagen y significa el número de puntos por pulgada de la imagen (dpi = dots per inch) (ppp = puntos por pulgada).

² A este procesamiento, cuando la resolución de una imagen es incrementada, se le llama *sobremuestreo*.

escanear a 600dpi en blanco y negro produce archivos intermedios de menor tamaño, pero escanear a 600dpi es mucho más lento en la mayoría de los escáneres. Además, resulta más fácil procesar escaneados en escala de grises a 300dpi porque tienen menos “basura digital” que los escaneados en blanco y negro a 600dpi.

- Es casi imposible mejorar la calidad de una imagen de libro obtenida con un escaneo pobre o procesada incorrectamente. Por ejemplo, algunos ebooks son hechos por gente inexperta a 150dpi, o en color en vez de blanco y negro. Los archivos de estos ebooks resultan de un tamaño enorme. La calidad visual e impresa de tales ebooks es mala y **¡no puede** mejorarse! Es importante (y no es difícil) hacer correctamente el escaneo de las imágenes, con eso se asegura una gran calidad en los ebooks resultantes. ¡Siga leyendo!

Un ebook escaneado de alta calidad debe ser de tamaño pequeño, con una gran apariencia en pantalla y en hojas impresas, además de tener textos localizables. Existen muchas maneras para lograr eBooks escaneados de alta calidad; todos los métodos implican una resolución de **600dpi**. En nuestro caso, los archivos finales quedarán en formato **DJVU**³ y generalmente ocuparán de 5KB a 10KB por página.

Claro que puede experimentar con otros programas que domine. Por ejemplo, algunas personas utilizan PhotoShop con plugins especiales, Book Restorer, Corel PhotoPaint, RasterID, también Matlab e IDLs para procesado de imágenes. Este tutorial ofrece un método concreto que prácticamente garantiza buenos resultados. Si es novato, por favor haga unos cuantos libros ajustándose a las instrucciones del tutorial. Verá cómo alcanza un alto nivel de calidad. Si luego desarrolla sus propios métodos, por ejemplo aplicar otras opciones del ScanKromsator o el uso de otros programas, será capaz de decidir cuál camino es mejor, puesto que ya podrá comparar la calidad de los otros resultados con la calidad de “referencia”, obtenida por el método sugerido en este tutorial.

Unas palabras de advertencia relativas al uso de FineReader para escanear. Por favor **¡no use** FineReader para escanear y procesar ebooks! El FineReader es un buen programa *para hacer OCR solamente*, pero no es óptimo para escanear y procesar imágenes con la intención de hacer un ebook escaneado. Fine Reader trata de ofrecerle una solución *todo-en-uno*, para escanear y procesar eBooks; resista caer en esta tentación, no use solo un programa para hacer todo. No obtendrá buenos resultados con FineReader; en todo caso, no pierde nada con seguir este tutorial.

³ Si no sabe qué es el formato DJVU, por favor use el Google o consulte Wikipedia para leer sobre ello. El formato DJVU fue desarrollado especialmente para almacenar imágenes escaneadas en archivos de alta compresión. El formato PDF fue planeado para documentos creados en un procesador de textos, es decir, para documentos *vectoriales* en lugar de documentos *escaneados*. Los eBooks en formato PDF ocupan mucho más espacio y lucen menos que los realizados en formato DJVU.

FineReader tiene los siguientes inconvenientes: 1) Algunas veces usa JPEG para compresión de imágenes. ¡Esto no es apropiado para textos en blanco y negro! 2) Almacena imágenes internamente como TIFFs blanco y negro de 300dpi y las auto-rotata. Blanco y negro de 300dpi es conveniente para el OCR, pero no es óptimo para escaneos digitales de eBooks. El algoritmo de auto-rotación es defectuoso y produce imperfecciones en la imagen (líneas “quebradas”). La auto-rotación está firmemente codificada en el FineReader 7.x, 8.x y no se puede deshabilitar.⁴ 3) Si escanea en escala de grises a 300dpi, que es el procedimiento recomendado aquí, FineReader realizará todas las operaciones a 300dpi, en vez de remuestrear a 600dpi. ScanKromsator primero remuestrea a 600dpi y posteriormente realiza el procesamiento. Por estas razones, los resultados del procesamiento hecho por FineReader siempre serán inferiores.

2. Escanear un libro

Tome un volumen grueso. Seguramente pensará que sólo a un loco puede ocurrírsele escanearlo página a página. ¡Si, tiene razón! Pero usted puede ser esa clase de loco capaz de escanear libros de cualquier tamaño, sin grandes incomodidades si organiza bien su trabajo.

⁴ Hasta ahora, sólo al FineReader 9 se le ha agregado una opción para desactivar esta auto-rotación. Sin embargo, FineReader 9 no puede ser usado (aún) para producir capas OCR destinadas a archivos DJVU.

Proof. First we will show how transition functions can be constructed for the given map (1.1.4) and vice versa. For the given point $x_0 \in M$ we choose n linearly independent columns of the matrix $P(x)$, and let $e(x)$ be $N \times n$ matrix formed by these columns. In the neighborhood U of the point x_0 we have $\text{rank } e(x) = n$. Similarly, let $g(x)$ be an $n \times N$ matrix formed by n rows linearly independent in U . Then the $n \times n$ matrix ge is nondegenerate, and we define $e' = (ge)^{-1}g$. It is clear that

$$e'e = 1, \quad ee' = P(x)$$

in U (here 1 denotes the identity $n \times n$ matrix). As M is compact, there exists a finite covering U_i and the matrices e_i, e'_i , defined in U_i , satisfying the above equalities

$$e'_i e_i = 1, \quad e_i e'_i = P(x) \quad (1.1.5)$$

Transition functions $f_{ij}(x)$ can be defined by

$$f_{ij}(x) = e'_i(x)e_j(x) \quad x \in U_i \cap U_j.$$

Conversely, given transition functions, we can define the projector $P(x)$ as a block matrix, with $n \times n$ -blocks $p_{ij}(x)$ as follows. Let $\{\rho_i(x)\}$ be a quadratic partition of unity, that is a collection of nonnegative functions, such that $\rho_i(x) \equiv 0$ outside U_i and

$$\sum_i \rho_i^2(x) \equiv 1. \quad (1.1.6)$$

Set

$$p_{ij} = \rho_i f_{ij} \rho_j, \quad i, j = 1, 2, \dots, N. \quad (1.1.7)$$

Then, using (1.1.1), (1.1.2), we get

$$\sum_j p_{ij} p_{jk} = \rho_i \sum_j (f_{ij} \rho_j^2 f_{jk}) \rho_k = \rho_i f_{ik} \sum_j \rho_j^2 \rho_k = p_{ik},$$

which means that the blocks p_{ij} give a projector matrix. The *frame* and the *coframe* matrices e_i and e'_i can be given as

$$e_i = \begin{pmatrix} \rho_{i1} f_{i1} \\ \rho_{i2} f_{i2} \\ \vdots \\ \rho_{iN} f_{iN} \end{pmatrix}; \quad e'_i = (\rho_{i1} f_{i1}, \dots, \rho_{iN} f_{iN}).$$

It is easy to verify the identities (1.1.5) for these matrices. This completes the proof of the first part of the theorem.

For two isomorphic bundles E, E' with transition functions f_{ij}, f'_{ij} define the projectors P, P' by (1.1.7). Using lemma 1.1.3, we may write

$$P' = A R, \quad P = R A, \quad (1.1.8)$$

Proof. First we will show how transition functions can be constructed for the given map (1.1.4) and vice versa. For the given point $x_0 \in M$ we choose n linearly independent columns of the matrix $P(x)$, and let $e(x)$ be $N \times n$ matrix formed by these columns. In the neighborhood U of the point x_0 we have $\text{rank } e(x) = n$. Similarly, let $g(x)$ be an $n \times N$ matrix formed by n rows linearly independent in U . Then the $n \times n$ matrix ge is nondegenerate, and we define $e' = (ge)^{-1}g$. It is clear that

$$e'e = 1, \quad ee' = P(x)$$

in U (here 1 denotes the identity $n \times n$ matrix). As M is compact, there exists a finite covering U_i and the matrices e_i, e'_i , defined in U_i , satisfying the above equalities

$$e'_i e_i = 1, \quad e_i e'_i = P(x) \quad (1.1.5)$$

Transition functions $f_{ij}(x)$ can be defined by

$$f_{ij}(x) = e'_i(x)e_j(x) \quad x \in U_i \cap U_j.$$

Conversely, given transition functions, we can define the projector $P(x)$ as a block matrix, with $n \times n$ -blocks $p_{ij}(x)$ as follows. Let $\{\rho_i(x)\}$ be a quadratic partition of unity, that is a collection of nonnegative functions, such that $\rho_i(x) \equiv 0$ outside U_i and

$$\sum_i \rho_i^2(x) \equiv 1. \quad (1.1.6)$$

Set

$$p_{ij} = \rho_i f_{ij} \rho_j, \quad i, j = 1, 2, \dots, N. \quad (1.1.7)$$

Then, using (1.1.1), (1.1.2), we get

$$\sum_j p_{ij} p_{jk} = \rho_i \sum_j (f_{ij} \rho_j^2 f_{jk}) \rho_k = \rho_i f_{ik} \sum_j \rho_j^2 \rho_k = p_{ik},$$

which means that the blocks p_{ij} give a projector matrix. The *frame* and the *coframe* matrices e_i and e'_i can be given as

$$e_i = \begin{pmatrix} \rho_{i1} f_{i1} \\ \rho_{i2} f_{i2} \\ \vdots \\ \rho_{iN} f_{iN} \end{pmatrix}; \quad e'_i = (\rho_{i1} f_{i1}, \dots, \rho_{iN} f_{iN}).$$

It is easy to verify the identities (1.1.5) for these matrices. This completes the proof of the first part of the theorem.

For two isomorphic bundles E, E' with transition functions f_{ij}, f'_{ij} define the projectors P, P' by (1.1.7). Using lemma 1.1.3, we may write

$$P' = A R, \quad P = R A, \quad (1.1.8)$$

Figura 1: Dos imágenes de la misma página, una hecha con una cámara digital y la otra con un escáner barato de cama plana. La imagen hecha por el escáner fue obtenida a 300dpi en escala de grises y remuestreada en blanco y negro a 600dpi. ¡A ver si adivina cuál es! Recomendamos que siempre use un escáner de cama plana y escala de grises a 300dpi o mayor resolución.

Primera nota: Por favor **¡no use** una cámara digital para escanear libros! Nunca obtendrá buenos resultados así sea una cámara carísima de 10 Megapíxeles o cualquier otra. Utilice un escáner ordinario de cama plana; incluso uno barato es apropiado. Lea el pie de la figura 1 y adivine cuál de las dos imágenes, obtenidas de la misma página, fue hecha con una cámara digital.

Para escanear, necesita cualquier programa que trabaje con el controlador TWAIN (Aplicación estándar destinada a la adquisición de imágenes con un escáner) del mismo escáner.⁵ Conviene que el programa sea capaz de guardar imágenes en el disco duro numerando los archivos como 0001.tif, 0002.tif o p0001.tif, p0002.tif, etc., por cada página escaneada. Por ejemplo, los visores de imágenes ACDsee, IrfanView o XnView también sirven para escanear imágenes y guardarlas así. También le podemos

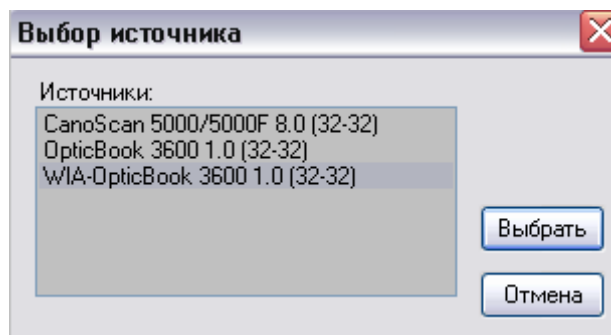
⁵ La mayoría de los escáneres están soportados por controladores TWAIN; para otros escáneres es probable que necesite controladores especiales.

recomendar que escanee con el programa VueScan, si su escáner lo reconoce y trabaja bien con él.

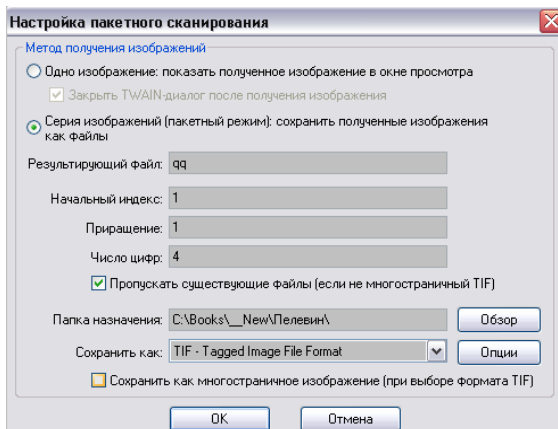
2.1 Configuración del IrfanView para escaneo

Como ejemplo, a continuación se describe cómo escanear con el IrfanView. (Este programa es gratuito.) El escaneo con otros programas es muy parecido.

Inicie el IrfanView. En el menú File (Archivo), pulse sobre “Choose TWAIN Source” (Seleccionar origen TWAIN) y elija el del escáner que vaya a usar.

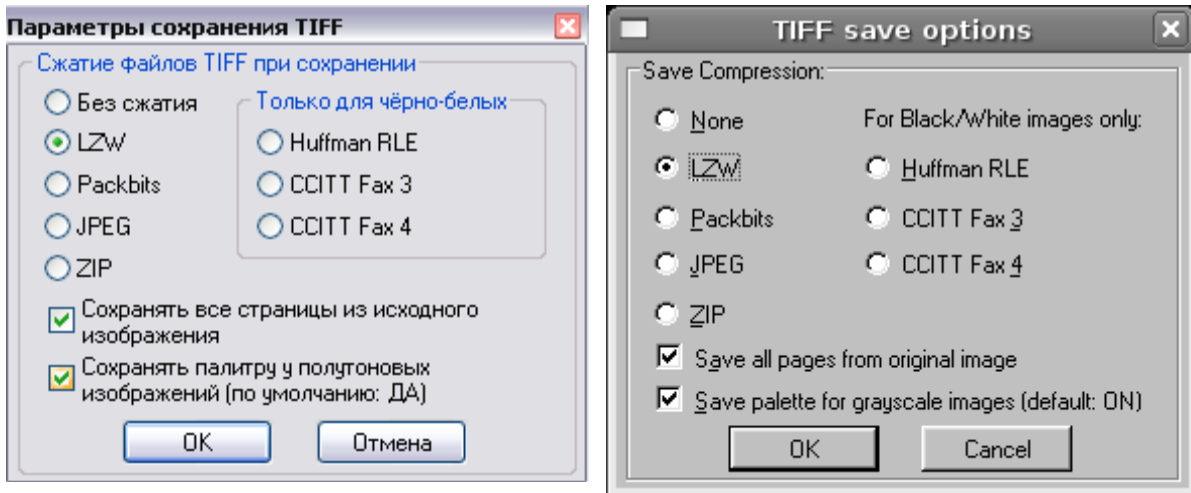


Luego en el mismo menú seleccione “Acquire/Batch scan” (capturar/lote escaneado).



Aquí deberá seleccionar la forma de numerar los archivos escaneados, dónde guardarlos y en qué formato hacerlo. Como se muestra, los archivos serán nombrados page0001.tif, page0002.tif, etc. Seleccionará **TIFF** como formato de imagen. (**¡No use** el formato de salida JPEG!).

Haga click en Options (Opciones), a la derecha del campo “Save as” (Guardar como). Allí se configuran las opciones para el formato TIFF.



Debe seleccionar la compresión LZW; así reduce el tamaño del archivo a la mitad, comparado con la opción sin compresión (“None”).⁶ Si posteriormente encuentra problemas de incompatibilidad con estos archivos TIFF (por ejemplo si después utiliza un programa que no puede abrirlos), entonces será necesario cambiar el método de compresión. Pero

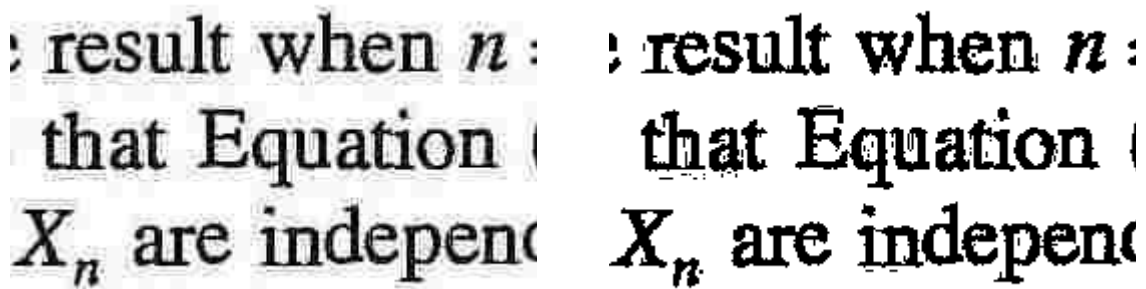


Figura 2: Aparecen desagradables sombras digitales debido a la compresión JPEG del blanco y negro. (En este ejemplo, la calidad configurada para la codificación JPEG fue muy baja, así estas sombras son bastante notorias a la vista). A la izquierda: imagen en escala de grises con sombras onduladas alrededor de las letras. Estas “sombras digitales” son típicas en la compresión JPEG de imágenes en blanco y negro. A la derecha: la misma imagen al convertirla nuevamente a blanco y negro, el resultado acumula “ruido digital”.

¡no use la compresión JPEG para textos en blanco y negro! La compresión JPEG provoca *alteraciones digitales*, estas son sombras desagradables

⁶ Típicamente, una página escaneada en escala de grises ocupará entre 2 y 4 megabytes del disco duro con la compresión LZW.

alrededor de las letras (vea la figura 2). Realmente carece de sentido utilizar JPEG para imágenes en blanco y negro.⁷

Ahora pulse sobre OK y pase a la ventana del controlador TWAIN de su escáner.

En la ventana TWAIN (puede ser otra ventana de configuración si no está usando controladores TWAIN), ajuste la resolución a **300dpi** y el modo de color a **greyscale** (escala de grises). Estas son las configuraciones más importantes.

2.2 Trabajo manual al momento de escanear

La labor realmente no es complicada:

- Para asegurarse que todo funciona bien, primero deberá hacer pruebas con alguna parte del libro. Tome el libro, ábralo donde las páginas tengan bastante texto y colóquelo sobre el cristal del escáner, con ambas páginas hacia abajo.
- Si fuera necesario presione con la mano, de manera que las páginas queden bien pegadas al cristal. (O ponga algo pesado encima del libro, como otro libro más grueso; aunque esto ejerce menos presión que la hecha con las manos.) **ADVERTENCIA:** demasiada presión frenaría el libre desplazamiento de la lámpara, incluso puede llegar a romper el cristal del escáner.
- Haga un “escaneo preliminar”. Así podrá examinar, en la ventana de vista previa, como quedará el escaneado. Si lo requiere, gire la página 90 grados para que el texto quede derecho. También en esta etapa ajuste el contraste, el brillo o la corrección gama si fuera necesario. El objetivo de esto es que el texto se vea con claridad.
- Con el ratón seleccione la región que capturará. Esta región debe incluir un espacio en blanco alrededor del bloque de texto.
- Pulse el botón “Scan” con el ratón y espere hasta que el escáner termine el recorrido de la página. Así obtiene el escaneado de una

⁷ Actualmente el formato JPEG no puede manejar imágenes en blanco y negro; cuando uno convierte imágenes blanco y negro en JPEG, el programa habrá de convertir las imágenes a escala de grises. Entonces la compresión JPEG introduce cierta pérdida de calidad, como se ve en la figura. La pérdida de calidad en la compresión JPEG es aceptable para fotografías, pero puede degradar significativamente los textos en blanco y negro, a menos que se seleccione un modo JPEG de alta calidad. (La calidad de la compresión JPEG es habitualmente seleccionable entre 1% y 100%. Las alteraciones se harán invisibles con calidades del 90% o mayores. Pero algunos programas, especialmente los usados para crear archivos PDF o para “optimizar” imágenes, no permiten que se configure manualmente la calidad de JPEG.)

página (o dos a la vez, si caben en la ventana del escáner). El archivo escaneado se guardará en el disco.

- Una vez configurado el programa de escaneo, proceda a escanear todas las páginas con esa misma configuración. Mientras la lámpara del escáner regresa, cambie a la siguiente página y ponga el libro otra vez en el mismo lugar del escáner. Pulse con el botón del ratón y vuelva a escanear. (El ratón puede estar a la izquierda apuntando sobre el botón “Scan”, así no necesitará estar volteando a ver continuamente. Algunos escáneres tienen botones físicos alternos y con alguno de ellos también será posible accionar el escaneo.)

Esta técnica le permite escanear libros enteros, página tras página, sin estar viendo constantemente la pantalla de la computadora o el teclado. Hasta podrá ver la TV o cualquier otra cosa mientras escanea. Según la velocidad de su escáner, puede lograr entre 100 y 200 escaneos por hora. Algunos escáneres son particularmente rápidos (como el Plustek OpticBook).

No es necesario colocar el libro sobre el cristal de manera que quede totalmente derecho (el borde del libro con el borde del escáner). Trate de ponerlo razonablemente derecho, pero es inevitable que algunas páginas escaneadas no queden derechas; es decir, algunas páginas quedarán *ligeramente torcidas*. Esta pequeña inclinación es tolerable y será corregida posteriormente (al finalizar el escaneo) mediante software. La corrección de esta inclinación se conoce como *deskewing* (enderezado).

Cuando escanee evite inclinaciones muy grandes así como páginas cortadas, en otras palabras, no deje texto fuera de la región de escaneo. Las arrugas en las zonas que rodean al texto a menudo dificultan el escaneo. En esos casos pruebe escanear una página a la vez (en vez de las dos páginas) o presione ligeramente más fuerte sobre la cubierta del libro. Es importante que el texto esté en contacto total con el cristal de escáner. Una hoja que quede a 1 mm de distancia del cristal hará que la imagen escaneada se vea muy borrosa ¡en casi cualquier escáner!

Resulta más rápido escanear un libro por pares de páginas, en vez de hacerlo una por una. Sin embargo, no todos los libros podrán escanearse de esta manera; algunos son muy grandes o no abren lo suficiente para escanearse de dos en dos páginas. Haga pruebas y determine cómo procederá. En todo caso, con el software de procesamiento podrá cortar las imágenes para dejarlas como páginas individuales.

En esta etapa, el resultado será un directorio lleno de archivos TIFF. Estos archivos son la *materia en bruto* que procesará al finalizar el escaneo. Considere que necesita tener suficiente espacio en el disco para almacenar

los escaneados (¡cuando menos 4 MB por imagen capturada!). Al terminar de escanear, use el modo de presentación de algún visor de imágenes, con él haga una rápida vista previa de las imágenes para asegurarse que no omitió alguna y que todas están escaneadas de manera correcta. Sería demasiado tarde descubrir en la etapa final del proceso que falta alguna página o que otras están de cabeza, ¡sobre todo cuando ya no tiene el libro con usted!

Nota: Cuando escanee un libro, ¡por favor no omita las páginas de títulos, la carátula, la información sobre la editorial, el índice general, el índice de materias, la bibliografía, las páginas en blanco, los números de página o cualquiera otro dato importante! No ahorrará mucho tiempo si decide no escanear estas 20 páginas más o menos. En cambio, un libro de ciencias es casi inútil sin su bibliografía, su índice o sin la información exacta acerca de su publicación. Tampoco piense que hará su vida más fácil, desde el punto de vista legal, si no escanea la información de la editorial. Lo que sí puede evitar es el escaneo de los sellos de la biblioteca (sólo cúbralos con una hoja, o quítelos después con el editor de imágenes), seguro que nadie quiere ver esos sellos en los ebooks.

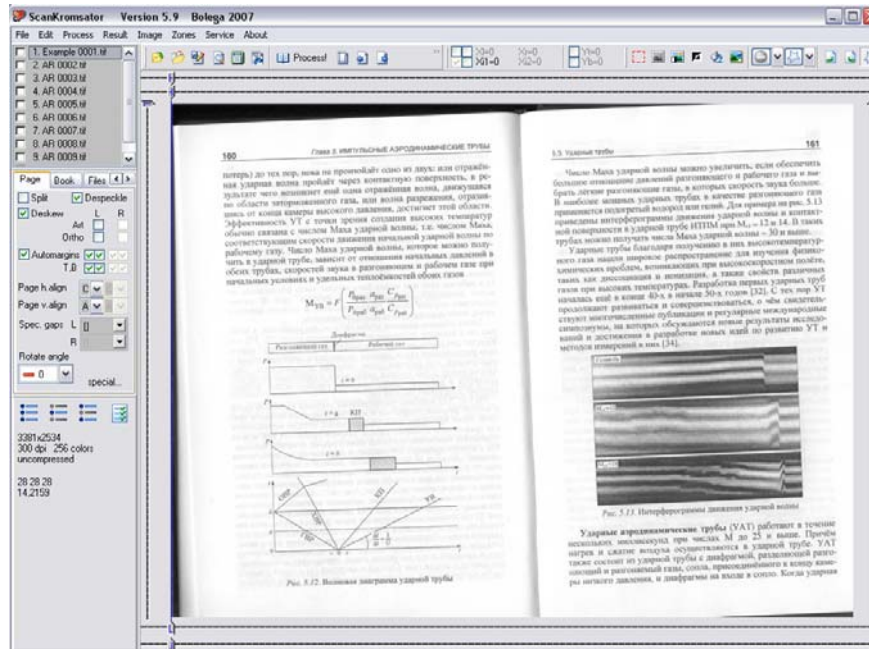
3. Procesar escaneados con el ScanKromsator

El actor principal en software para procesamiento es el asombroso ScanKromsator, desarrollado por Bolega.⁸ ScanKromsator es una herramienta muy poderosa para procesar material escaneado. Tiene un gran número de valiosas funciones, pero algunas de ellas no son intuitivas o son difíciles de entender si sólo se le da un vistazo a su interfaz de usuario.⁹ En este tutorial recorrerá una secuencia simplificada de trabajo con el ScanKromsator, se supone que ya tiene escaneado un libro en escala de grises a 300dpi.

Inicie el ScanKromsator y cargue los archivos TIFF en bruto (hágalo desde el menú File). La lista de archivos aparecerá en la columna arriba a la izquierda. La **barra de herramientas** con algunas **pestañas** (“Page”, “Book”, etc.) aparecerá debajo de la lista de archivos.

⁸ Por favor no envíe correos a Bolega para solicitarle ayuda, documentación, código fuente del ScanKromsator o la adición de nuevas características. En lugar de eso, ¡aprenda a usar bien el programa y a crear eBooks de buena calidad!

⁹ Aquí sólo hablaremos de las funciones del ScanKromsator en forma muy reducida. Desafortunadamente este programa aún no cuenta con un manual de usuario que describa ampliamente todas sus funciones.

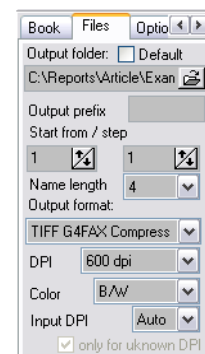


En el ejemplo, se muestra un libro que fue escaneado a dos páginas por pantalla y, como se ve, presenta una ligera inclinación. Ahora nuestra labor con esa imagen será separar las páginas, enderezarlas y cortarlas de manera que cada página quede del mismo tamaño y con márgenes iguales. Si su escaneado es de una sola página, ya no necesitará separar, pero sí tendrá que enderezarla y cortarla. A esta operación el programa le llama “kromsating”.¹⁰

3.1 Ejecución en borrador

El primer paso es realizar el proceso en un *borrador*, es decir, hacer el anteproyecto o la preparación para el proceso final de los archivos en bruto.

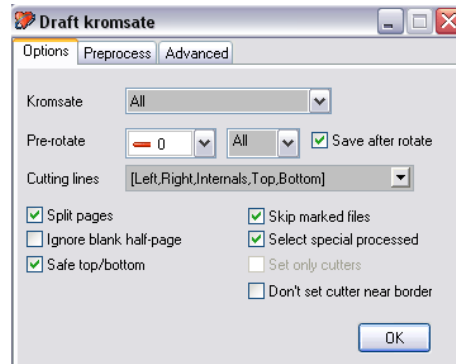
Haga clic en la pestaña “Files” de la barra de herramientas. Surgirá una ventana de diálogo donde podrá configurar la resolución de salida (¡esto es muy importante!) a 600dpi, la carpeta para almacenar los archivos de salida (esta *Output folder* tiene predeterminado el subdirectorio out del directorio actual), y la forma de numerar dichos archivos (prefijo *Output prefix*, número de dígitos *Name length*, número inicial *Start from*, secuencia *Step*). Observe que el formato para archivos comprimidos es: codificación TIFF



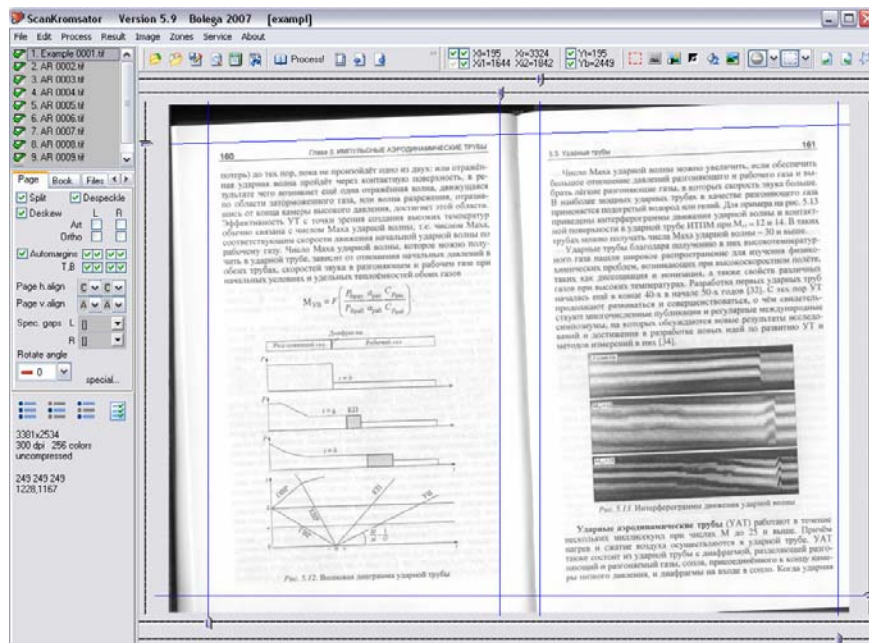
¹⁰ La pseudo palabra “kromsate” es la distorsión de un término ruso y da la idea de “cortar en piezas o trozos”. En el ScanKromsator, “kromsate” se refiere al proceso de dividir un escaneado de doble página en imágenes de páginas individuales, así como al recorte de estas páginas de manera que sus márgenes queden nivelados e iguales en todas ellas.

G4, el cual es óptimo para imágenes en blanco y negro. Este será el formato de salida después del proceso.

Para desarrollar el anteproyecto haga clic en el botón “Draft kromsate” (Corte en borrador); es el icono etiquetado con unas tijeras, se localiza a la izquierda del botón “Process” de la barra de herramientas. Cuando pulse en el botón “Draft kromsate” obtendrá la ventana mostrada a la derecha. En esta ventana deberá marcar en “Split pages” (Dividir páginas) y “Safe top/bottom” (Proteger arriba/abajo). En el campo “Kromsate” la opción “All” significa que las opciones marcadas se aplicarán a todas las páginas. Si algunas páginas no necesitan dividirse, entonces elija “Current” (Actual) en ese campo y desmarque “Split pages” para esas páginas.



Pulse sobre OK y espere de 10 a 15 minutos hasta que la operación “Draft kromsate” termine. Se mostrará una pantalla como la de la figura.



Observe que ahora hay “palomitas” verdes en la lista de páginas (arriba a la izquierda de la columna), significa que esas páginas han pasado por el “corte en borrador” satisfactoriamente. En cada página verá líneas azules que la cruzan. Estas líneas son las *marcas de corte* que determinan como será cortada y dividida la imagen de página. Note que el programa intenta determinar automáticamente donde cortar los márgenes y dónde dividir la doble página en páginas individuales. En algunos casos el programa puede

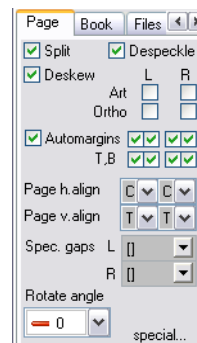
cometer un error y cortar en exceso o muy poco; en esos casos usted después podrá ajustar manualmente dichas posiciones de corte.

3.2 Configuración de las opciones

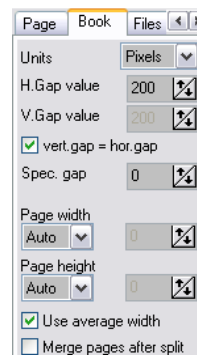
El siguiente paso importante es ir a las opciones de procesamiento y prepararlas para que el ScanKromsator haga la ejecución definitiva (ya no en borrador). Las opciones de procesamiento se ajustan en las distintas pestañas de la barra de herramientas (a la mitad de la columna izquierda).

Por favor observe: Cada opción se puede configurar ya sea para aplicarse a todas las páginas a la vez o sólo a la mostrada en pantalla. Para aplicar una opción a todas las páginas, mantenga pulsada la tecla **Ctrl** mientras hace clic con el ratón en el cuadro de selección de la opción. De esta manera, ajusta rápidamente las opciones comunes para todo el trabajo y luego vaya a las páginas problemáticas para destinarles individualmente otras opciones.

Primero haga clic en la pestaña “Page” (Página). Aquí se configuran las opciones del proceso de corte de las páginas. La opción “Split” (Dividir) significa separar las imágenes de doble página en páginas individuales. “Deskew” (Enderezar) enderezará las imágenes de cada página por separado. “Despeckle” (Desmanchar) quita todas las manchitas. Algunas veces “Deskew” produce páginas bastante inclinadas; se debe generalmente a algunas ilustraciones complicadas. En estos casos, marque la opción “Art” para esas páginas. Puede configurar “Ortho” (Recto) si la página necesita rotarse 90 grados. Puede ajustar estas opciones de forma separada para páginas izquierdas y derechas (L y R).

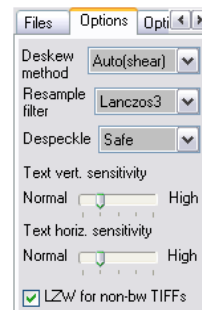


Luego haga clic en la pestaña “Book”. Aquí se configura todo lo relativo al tamaño y distribución de las páginas para el libro terminado. “H.Gap” (Distancia horizontal) es el tamaño de los márgenes horizontales. Un valor de 200 es bueno para 600dpi (equivale a 1/3 de pulgada). La anchura y altura de las páginas pueden dejarse en Auto. También aquí puede colocar las páginas en distintas posiciones (alineada al centro/alineada arriba/alineada abajo).

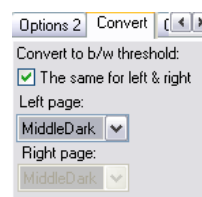


Ya vio el contenido de la pestaña “Files” (Archivos) en la etapa “borrador”. Recuerde que ¡es **muy importante** mantener 600dpi como resolución de salida en el menú “Files”!

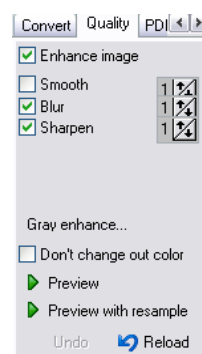
Ahora, desplácese en la barra de pestañas y haga clic en “Options” (Opciones). Ponga la configuración de “Deskew method” (Método de enderezado) en Auto(shear, *recortar*), el “Resample filter” (Filtro de remuestreo) en Lanczos3. La configuración de “Despeckle” (Desmanchar) puesta en Fine+Normal (Fina + Normal) o en Safe (Segura) lleva a un método de desmanchado “inteligente”, esto evita quitar los puntos obre la i o la j, por ejemplo. Los controles “Text sensitivity” (Sensibilidad del texto) fijan el nivel lógico del auto corte. Una baja sensibilidad podría cortar los números de página si se encuentran bastante alejados del texto. Talvez necesite ajustar un poco la configuración de sensibilidad; pero en la mayoría de los casos no es necesario cambiar este ajuste.



Sáltese por ahora la pestaña “Options 2” (Opciones 2) y haga clic en la pestaña “Convert” (Convertir). Aquí se configura el umbral de conversión de imágenes en escala de grises a imágenes blanco y negro. No olvide mantener pulsada la tecla Ctrl (para aplicar el ajuste a todas las páginas), y elija para “Threshold” (Umbral) la opción MiddleDark (semi oscura). Experimente con otras configuraciones si no le agradan los resultados.

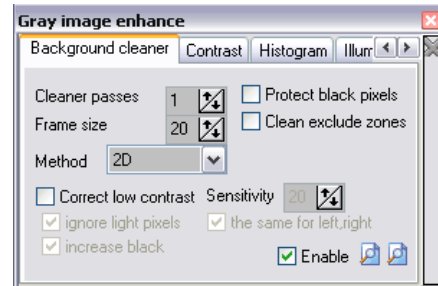


Haga clic en la pestaña “Quality”; allí puede controlar mejor la conversión a blanco y negro. ¡Esta función es muy importante! Marque “Enhance image” (Mejorar la imagen), “Blur” (Borroso) póngalo en 1 y Sharpen (Nitidez) también en 1. Lo importante con esta configuración es que la imagen quede lo mejor suavizada posible. Los valores de Blur y Sharpen podrían ser 2 en vez de 1, aunque el valor de 1 generalmente es bueno. Un valor muy grande hará que las letras sean más negras. Talvez necesite experimentar, según la calidad de impresión obtenida con algún libro en particular.

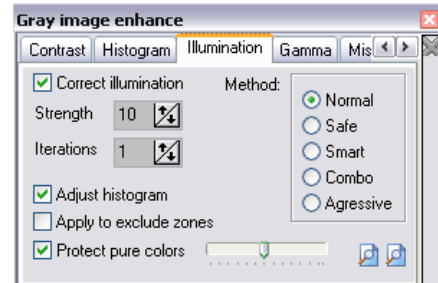


Otra función importante es “Gray enhance” (Mejorar grises). Haga clic en esta opción, puesto que tiene imágenes de escaneadas en escala de grises (¡así debe ser!).

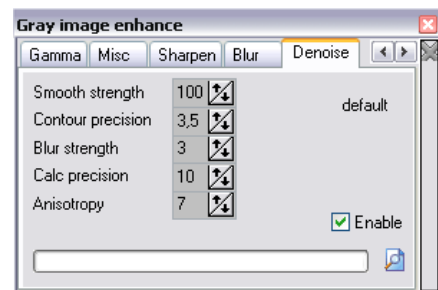
Obtendrá una ventana de diálogo con varias opciones para el manejo de imágenes en escala de grises. Vaya a la pestaña “Background cleaner” (Limpiar fondo) y marque “Enable” (Permitir).



Sáltese algunas pestañas y haga clic en la de “Illumination” (Iluminación); haga clic en “Correct illumination” (Corregir iluminación). Esta condición normalizará (equilibrará) la iluminación de toda la página, lo cual es importante ya que generalmente algunas zonas de la misma quedan más oscuras que otras. Esta es una característica muy útil para eliminar las sombras negras, que de otra forma podrían aparecer en los lugares más oscuros de la página.



Sáltese otras pestañas y haga clic en “Denoise” (Eliminar ruido). Configure los parámetros como se muestra en la figura de la derecha. Estos parámetros limpian la imagen. Estas son las últimas opciones que ajustaremos para dejar de molestar ya.

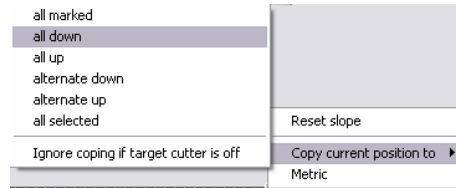


Puede usar el menú para entrar a File / Options... (Archivo/Opciones...) y guardar las opciones en un archivo. Esto le ahorrará trabajo en las próximas ocasiones que use el programa.

El último paso, antes de pasar al procesado principal, es una revisión visual de la posición de los cortadores. Necesitará ir *página por página* para verificar que los cortadores están colocados correctamente. Sí, esto es bastante aburrido... pero se hace rápido.

Ponga dos dedos de su mano izquierda sobre las teclas **q** y **w**; pulse estas teclas para regresar o avanzar las páginas. Con la mano derecha, sostenga el ratón y ajuste la posición de los cortadores según se requiera. Algunas veces es por una sombra inclinada, o por alguna otra razón, que debemos poner la línea de corte *en un ángulo* fuera de la vertical u horizontal. Para lograrlo mantenga pulsada la tecla **Shift** y arrastre la línea de corte por un extremo.

También puede copiar la posición del cortador de una página a otra. Coloque el cursor sobre el cortador y haga clic en el botón derecho del ratón, verá el menú que se muestra a la derecha. Por ejemplo, si la posición actual del cortador debe ser aplicada a todas las páginas subsecuentes, haga clic en “Copy current position to” (Copiar posición actual a) / “all down” (todas las que siguen).



Si alguna página contiene fotografías o figuras en color, necesitará protegerla de la conversión a blanco y negro. Esto puede hacerse cuando se verifican las posiciones de corte, consiste básicamente en seleccionar de forma arbitraria alguna parte de la página y marcarla como *picture zone* (zona de ilustración). Vea la Sección 4 para mayores detalles.

Guarde ahora las configuraciones para este trabajo mediante la orden File/Save Task (Archivo/Guardar trabajo), que está en el menú principal. Esta opción es útil si quiere hacer una pausa en el trabajo y continuarlo después.

3.3 Ejecución definitiva

Ahora que todo está listo, comience la ejecución principal del ScanKromsator. Vaya a la parte superior en la barra de herramientas y pulse en el botón largo que dice “Process!” (Procesar), se muestra con el icono de un libro:



El programa le preguntará si confirma que realmente quiere cambiar la resolución de las imágenes. ¡Confirme! El proceso comenzará ahora.

Habrá de esperar un rato. La operación de sobremuestreo puede ser bastante lenta; en las versiones recientes del ScanKromsator (5.8 y superiores) esta operación se hace más rápido. Se espera un procesado de 5 páginas por minuto más o menos. Cuando todo haya terminado, deben verse los archivos de salida en la carpeta de salidas. Verifique que todas las páginas fueron cortadas y enderezadas correctamente. Si encuentra páginas que no se procesaron bien, repita el proceso sólo en esas páginas aplicando algunas otras opciones.

La ejecución del proceso principal llega a tomar algunas horas en computadoras lentas. Aquí cabe observar que no es necesario procesar el libro completo en una sola ejecución. Pero tenga en mente que si procesa

un nuevo lote de páginas; entonces deberá configurar Book/Page width/Fixed (Libro/Ancho de página/Fijo) al tamaño que estableció para las páginas del lote anterior (de manera que todas las páginas tengan las mismas dimensiones al finaliza el proceso). Para determinar el tamaño de las páginas, es suficiente tomar de 10 a 15 páginas y probar.

Si prefiere, puede usar las potentes características de limpieza del ScanKromsator para eliminar la “basura digital” de algunas páginas. Se le llama “basura digital” a cualquier mancha extraña en el papel como marcas de tinta o lápiz, sellos de biblioteca, etc. En caso contrario, también puede usar cualquier editor de gráficos para limpiar manualmente las imágenes. Con un poco de suerte, talvez no sean muchas páginas que limpiar.

4. Procesar figuras en color y fotografías

Hablaremos por separado de las figuras en color porque, en los viejos libros de ciencias, su uso no fue muy frecuente. No obstante, su lugar a lo largo del proceso está en la etapa donde se verifica y ajusta la posición de los cortadores.

En la versión más reciente del Kromsator (la 5.9) se incluye una característica para el procesado de imágenes en color, se le ha llamado *picture zones* (zona de ilustraciones). Algunas páginas de los libros podrán ser una *ilustración*, por ejemplo cualquier imagen que no esté en blanco y negro, como una fotografía o un diagrama a colores. Necesitará proteger estas imágenes de la conversión a blanco y negro. Marque la zona de la ilustración, arrastre el ratón sobre la ilustración para dejarla dentro de un rectángulo de selección y luego haga clic en el botón “Mark as picture zone” (Separar como zona de ilustración), el icono relacionado está

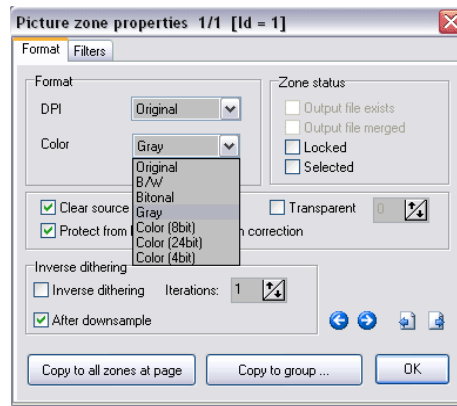
enmarcado en azul en la barra de herramientas:



También existe la posibilidad de crear un contorno poligonal alrededor de una zona de ilustración. Esto es útil, por ejemplo, cuando la página fue escaneada con mucha inclinación. Utilice el botón de herramienta *star-shaped* (contorno en estrella) para marcar tales zonas:



Para configurar una zona de ilustración, haga doble clic sobre la región seleccionada. Aparecerá la ventana de diálogo “Picture zone properties” (Propiedades de la zona de ilustración).



Aquí necesita configurar el color de la ilustración. Por ejemplo, si la página contiene una fotografía en escala de grises (en lugar de una fotografía o un diagrama en color), elija Color = Gray (Color = Gris).

No hablaremos aquí de otras opciones de las zonas de ilustración; como ve, hay muchas opciones más destinadas a los usuarios avanzados. Observe que después pasar imágenes por la zona de ilustración del Kromsator, estas páginas se guardan en *archivos separados*. De manera que al finalizar el proceso principal tendrá que mezclarlos con los otros archivos de páginas. Esta operación se realiza desde el menú principal, con las órdenes *Zones/Picture zone/Merge zones* (Zonas/Zona de ilustración/Mezclar zonas). Los archivos resultantes de esas páginas estarán en formato TIFF, el texto aparece en blanco y negro pero las ilustraciones conservan sus colores.

5. Codificar escaneados en DJVU

Una vez finalizado el proceso de los escaneados en bruto, tendrá en la carpeta de salida un grupo de archivos TIFF, todos (o casi todos) en blanco y negro a 600dpi. Estos archivos TIFF ocupan típicamente entre 50 y 100 KB por página, en lugar de los 4 MB que ocupan los que aún están en escala de grises. Por ahora deberá revisar estos archivos TIFF para asegurarse que la calidad de las imágenes en blanco y negro es buena: que las letras estén bien definidas, que tengan contornos suaves, que haya poca o ninguna basura digital, etc. Para la revisión de todo eso, abra los archivos TIFF en un visor de imágenes (tal como el IrfanView), con alto grado de acercamiento (bastante zoom).

Aún, 50 a 100 KB por página es demasiado. El siguiente paso es convertir esas imágenes al formato DJVU; esta codificación reducirá drásticamente su tamaño, típicamente quedarán páginas de entre 5 y 10 KB.

Para hacer un buen, o bastante aceptable, archivo DJVU, necesita cualquiera de estos dos programas: el DjvuSolo versión 3.1 o el Djvu Document Express (DDE) 4.x, 5.x, 6.x (o su opción completa Djvu Document Express Enterprise (DEE) versión 5.1, 4.x, 5.x, 6.x).¹¹ Los programas DDE y DEE son mucho más rápidos que el DjvuSolo, y el DEE 5.1 es capaz de configurarse para el proceso por lotes. A cambio, DjvuSolo es un programa pequeño y gratuito. En términos de los resultados obtenidos, la calidad de los archivos DjvuSolo lucen casi igual a los del DDE/DEE si sus opciones se configuran correctamente.

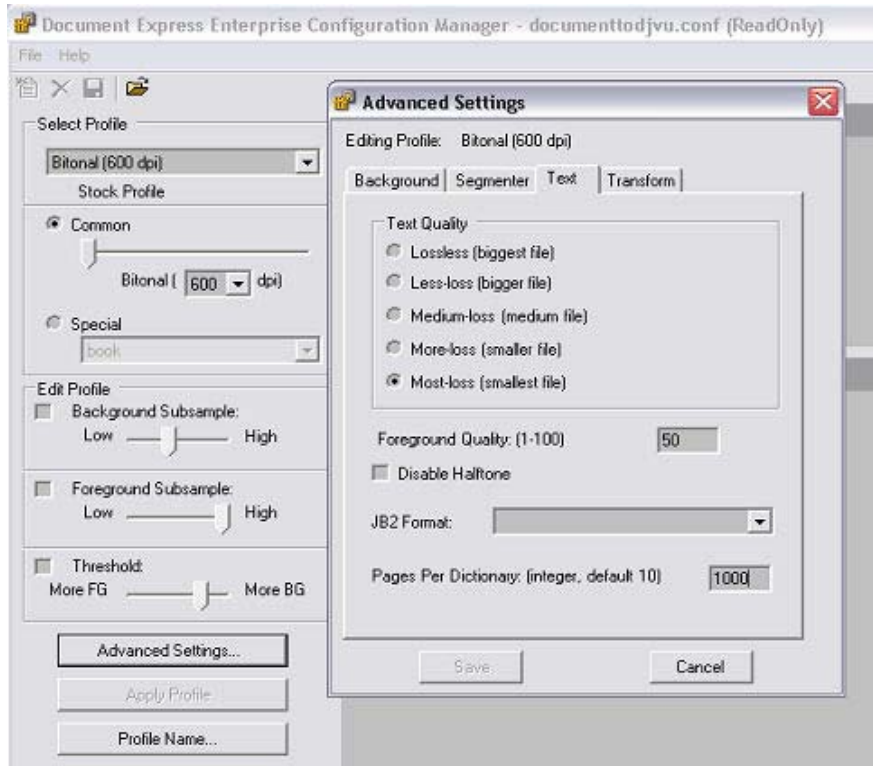
Existen dos maneras de crear archivos DJVU: una es manualmente y otra en un proceso por lotes. Para realizar archivos DJVU manualmente, ejecute DjvuSolo o DDE y haga clic en File/Open (Archivo/Abrir) para abrir el primer archivo TIFF. Ahora haga clic en Edit/Insert pages... (Editar/Insertar páginas...) y seleccione los archivos TIFF restantes. Por favor considere: la ventana de selección muestra error en este paso si eligió muchos archivos con la acostumbrada combinación de la tecla **Shift** y clic del ratón, los archivos sí aparecerán en la ventana pero colocados en orden inverso. Compruebe que los archivos seleccionados están en el orden correcto. Ahora guarde con "Save as..." (Guardar como...), en esa ventana elija para "Bundled" (Paquete) el formato DJVU y en "Bitonal" la opción a 600dpi. También podrá editar el archivo documenttodjvu.conf en el directorio perfiles (profiles) y configurar pages-per-dict (páginas por diccionario) = 100 ó 200. A mayor número de páginas por diccionario más lento será el proceso de compresión, pero el resultado será un archivo de menor tamaño.

Considere que la opción "Bitonal" (o perfiles) en los codificadores DJVU se refiere únicamente a escaneados en blanco y negro, mientras que la opción "Scanned" (Escaneado) se aplica a los escaneados que tienen algunos (no tantos) colores, sin llegar a las fotografías. Utilice la opción "Photo" (Fotografía) para imágenes fotográficas.

Para crear un archivo DJVU por lotes, se necesita el DEE 5.1.¹² Primero deberá elaborar un juego de opciones especiales (o "custom profile", perfiles de usuario), para realizar el trabajo de codificación DJVU. Para hacer esto ejecute el Document Express Configuración Manager (Administrador de configuración del Document Express), elija el perfil "Bitonal (600dpi)" de la lista de perfiles que se encuentra en "Select Profile" (Seleccione un perfil), haga clic en "Advanced settings" (Configuraciones avanzadas) y obtendrá la ventana de diálogo mostrada a continuación.

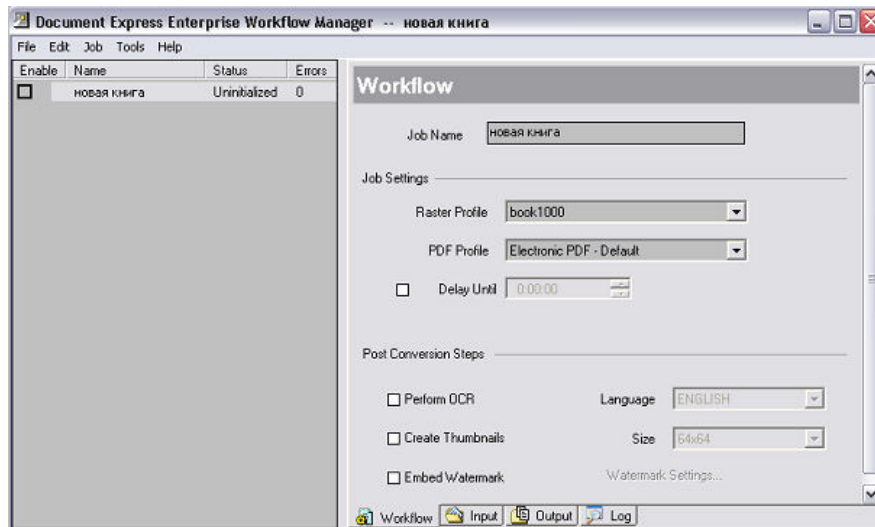
¹¹ Existe también un paquete de software libre llamado "DjVuLibre", pero no logra producir archivos DJVU bien comprimidos.

¹² Este es un paquete bastante grande; pero existe una versión desensamblada que ocupa sólo unos 20 MB de disco duro.



Entre a la pestaña “Text” (Texto), como se muestra arriba. En esta sección, configure “Pages per dictionary” (Páginas por diccionario) = 1000 (si esto agota la RAM de su computadora, o la hace muy lenta, baje a 200 o 300 en lugar de los 1000). Guarde este nuevo perfil de usuario con un nuevo nombre, digamos Bitonal-1. Haga lo mismo para el perfil “Scanned (600dpi)”, si va a codificar libros con dibujos en color.

Ahora ejecute el Document Express Workflow Manager (Administrador de trabajo del Document Express). Cargue todas las páginas TIFF en él. En el campo “Job name” (Nombre del trabajo), escriba el nombre que desee darle al libro. Elija el perfil de usuario previamente creado, de la lista que se mostrará en “Raster profile” (Perfil del tramado).



Luego haga clic en la pestaña “Output” (Salida), localizada en la barra inferior de la ventana. En la lista de “Separate document(s)” (Documentos separados), elija “One document only” (Un documento solamente). Marque la opción “Enable” (Habilitar) del extremo izquierdo. Espere a que finalice la codificación. Mientras tanto puede ver en la pestaña “Log” cuál es el progreso de la codificación. Eso es todo; el archivo DJVU está creado.

¡No borre aún los archivos TIFF! Podría tener que hacer nuevamente la codificación si el archivo DJVU tuviese algún error. Estos archivos TIFF también le servirán si tiene planeado aplicar un OCR (reconocimiento óptico de caracteres), como se explica en la sección 6.

El resultado de la codificación es un archivo multipágina DJVU que contiene el libro completo. Renombre el archivo para que muestre un nombre sensato; no lo deje sólo como math1.djvu. Muchas veces conviene que el nombre del archivo lleve el título del libro y su autor, tal vez el año de su publicación o su número de registro ISBN. Esto implica un trabajo insignificante, pero es más fácil compartir el archivo por Internet si tiene un nombre significativo.

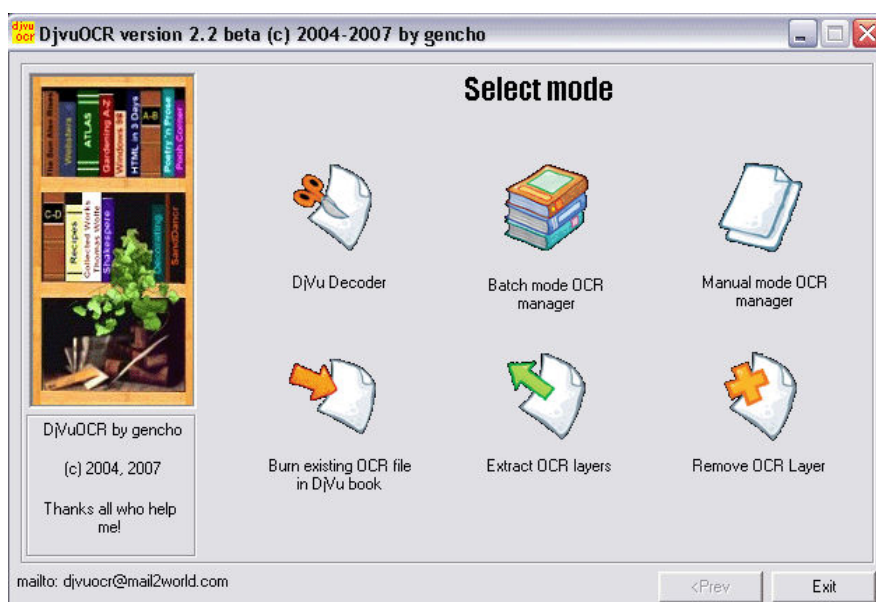
6. Crear capas de texto con OCR

Un eBook con buscador incluido es mucho más cómodo de usar. Comparado con el agobio de escanear y procesar el libro para crear el DJVU, realmente es una insignificancia hacerle un OCR al archivo.

La búsqueda dentro de los archivos DJVU funciona solamente si el archivo tiene la llamada capa OCR. Esta capa es básicamente una lista de palabras almacenada, en el mismo archivo DJVU, de manera comprimida. Para crear la capa OCR se usan estos dos programas: FineReader y

DjvuOCR. Necesita la versión 7 u 8 del FineReader, es suficiente con una muestra de evaluación o una versión no registrada, que logre descargar gratis.¹³ El resultado de ejecutar FineReader será un conjunto de archivos en un lote de FineReader. El maravilloso programa DjvuOCR, creado por Gencho, leerá directamente estos archivos, extraerá la información OCR y la insertará en los archivos DJVU.

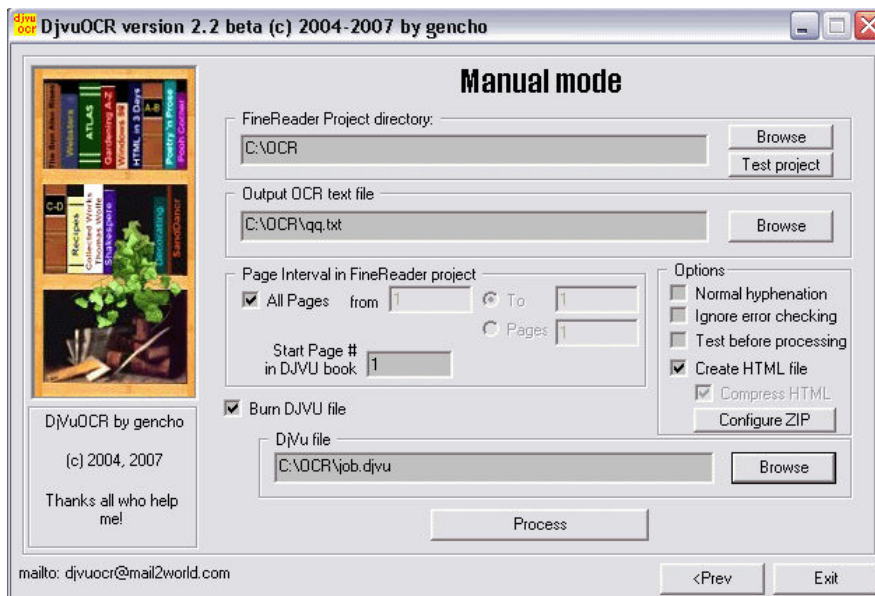
Se supone que ya creó el archivo DJVU a partir de los archivos TIFF y que no ha borrado estos últimos. Cargue los archivos TIFF en un lote nuevo con el FineReader (¡tenga presente los problemas que se dan al seleccionar muchos archivos a la vez!). Seleccione el idioma de reconocimiento y pulse en “Read all” (Leer todo). Cuando finalice el proceso OCR, haga clic en “Save batch” (Guardar lote). No se recomienda editar el texto OCR, ya que las versiones anteriores del DjvuOCR no podrían procesar los lotes de FineReader si el texto OCR ha sido editado. La versión reciente DjvuOCR 2.2, ya acepta ediciones mínimas. No debe reescribir grandes bloques de texto; en otras palabras, deberá conservar muchos símbolos originales en sus posiciones originales si hace alguna edición. Tampoco debe borrar los signos de fin de línea, de manera que el número de líneas de un párrafo tiene que conservarse. Por eso mejor le recomendamos que no edite nada del texto OCR. Después de guardar el lote FineReader, salga del programa y ejecute ahora el DjvuOCR.



Como puede ver, en la imagen de arriba, el programa tiene varias funciones; por ejemplo, “DjVu Decoder” (Decodificador Djvu) produce archivos TIFF a partir de uno DJVU, muy útil en caso de haber borrado sus archivos TIFF, o si trabaja con algún otro archivo DJVU. Por ahora,

¹³ FineReader 9 ya está disponible, pero esta versión no puede agregar OCR a los archivos DJVU y tampoco el programa DjvuOCR le da soporte.

sólo usará la función “Manual mode OCR manager” (Administrador OCR en modo manual). Haga clic allí y surgirá la siguiente ventana.



Seleccione el directorio donde guardó el lote FineReader, está en el campo “FineReader Project directory” (Directorio del proyecto FineReader). En “Output OCR text file” (Salida del archivo de texto OCR) ponga el nombre para el nuevo archivo; no necesariamente será el nombre mostrado. Marque la opción “Burn DJVU file” (Grabar en archivo DJVU) y elija en la ventana el archivo DJVU correspondiente; esto significa que los datos OCR serán adosados (“cocinados”) al archivo DJVU. Haga clic en el botón “Process” (Procesar), espere unos minutos a que termine y eso es todo. Ahora tiene un archivo DJVU ¡totalmente apto para la búsqueda de textos!

7. Agregar cubiertas y láminas de color al libro

Es razonablemente fácil agregar una cubierta sencilla al libro. Sólo escanee la cubierta del libro a 300dpi en color, o aún a 200dpi. Afine ligeramente la imagen en un editor de gráficos. Codifíquela a DJVU con el perfil “Photo(300)” o con “Scanned”. Resulta un archivo de página en DJVU que necesita ser insertado al comienzo del eBook DJVU, después de finalizar los otros procesos. Normalmente la cubierta del libro no es mayor de 20 o 30 KB. Seguramente no será necesario derrochar mucho esfuerzo en hacer una cubierta muy bonita. Considere que quienes lean su eBook pasarán más tiempo en los textos que admirando la cubierta.

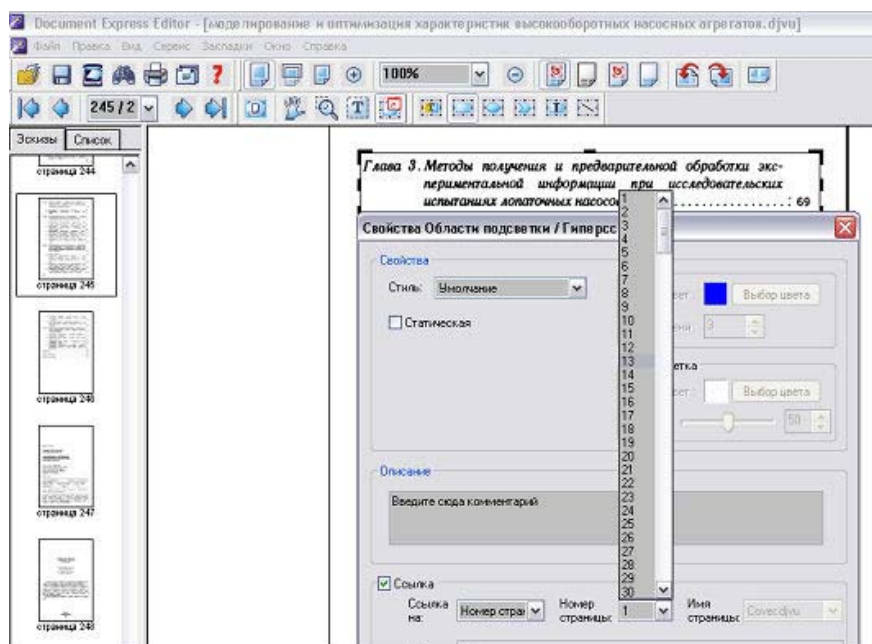
De la misma forma puede agregar láminas en color, esto es, páginas especiales que tienen sólo ilustraciones en color. Escanéelas por separado e insértelas en el DJVU terminado, después de hacer los otros procesos.

Para insertar o reorganizar páginas en un archivo DJVU, use el DjvuSolo o el DDE. Abra el archivo DJVU y verá las miniaturas de todas las páginas en la columna de la izquierda. Simplemente arrastre allí esas miniaturas a las nuevas posiciones para reorganizarlas; también puede “Cut” (Cortar), “Copy” (Copiar) y “Paste” (Pegar) páginas o grupos de páginas seleccionadas, o eliminar páginas. Para insertar más páginas a un archivo DJVU existente, vaya al menú Edit/Insert pages... (Editar/Insertar páginas...). Puede insertar archivos DJVU de una o más páginas (antes o después de cualquier página), según sus necesidades.

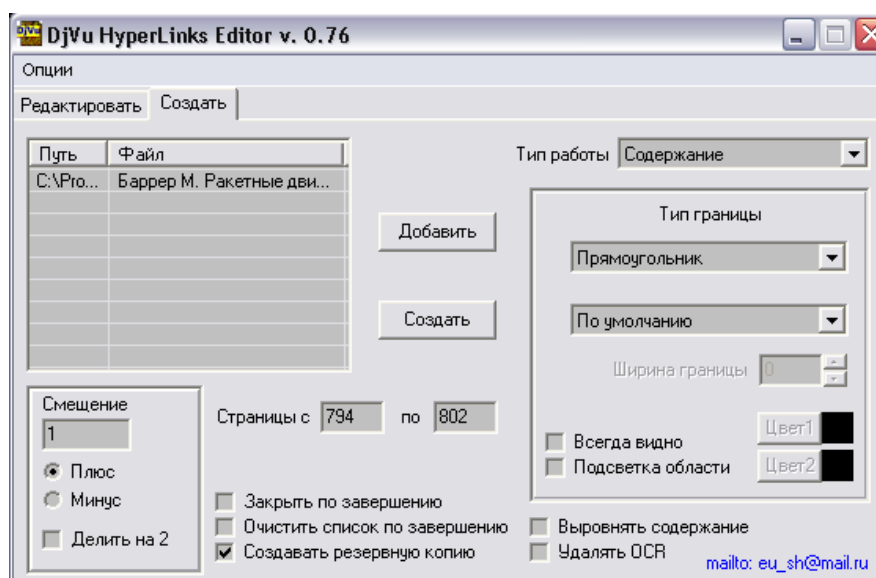
8. Agregar hiperenlaces y marcadores

Al finalizar completamente el trabajo con el archivo DJVU (incluso el OCR), ya podrá agregar hiperenlaces de navegación a su eBook. Hay dos formas de hacerlo.

La primera es con los programas DjvuSolo o Djvu Editor, para agregar los hiperenlaces a mano. Normalmente uno agrega los hiperenlaces en el índice, para facilitar la navegación del libro. En el DjvuSolo o Djvu Editor seleccione cualquier área rectangular de cualquier página y allí podrá insertar un hiperenlace hacia otra página del archivo DJVU. El usuario pasará a esa página cuando haga clic en alguna parte de la zona antes seleccionada. Considere que los hiperenlaces se dirigen a un número de página, así que para agregarlos debe hacerlo *después* de todo cambio en el orden de las páginas, o *después* de insertar cualquier página adicional al archivo DJVU. Podrá hacer tantas áreas rectangulares como desee y ponerles hiperenlaces hasta que el rostro se le ilumine de azul.



La segunda manera de agregar hiperenlaces es semi-utomática, para ello se usa el programa DJVU Hyperlinks Editor.¹⁴ Ejecute ese programa y verá la siguiente ventana.

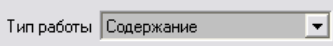


Primero debe definir las opciones para los hiperenlaces, así que especifique el rango de páginas () en el que está localizado el índice general en el archivo DJVU. Estos son los números de página DJVU, los cuales pueden diferir de los números impresos en las páginas del libro y en el índice general (esto se debe a la inclusión como páginas de las cubiertas y las páginas preliminares). Para corregir esa diferencia, generalmente agregaremos cierta compensación a la numeración de páginas; por ejemplo, la página 10 en el impreso puede ser realmente la 11 en el archivo DJVU, porque la cubierta es considerada como una página.¹⁵ Entonces necesita introducir la compensación correspondiente en el

campo (“Offset”) (Compensación). Ahora que se han introducido todas las opciones, pulse sobre el botón (“Add”) (Agregar). Esto agregará un nuevo archivo DJVU a la lista del panel izquierdo; las opciones actuales se aplicarán a ese archivo. Puede configurar otras opciones y agregarlas a diferentes archivos. Finalmente pulse en el botón (“Create”) (Crear). Esta acción insertará la información de los hiperenlaces en todos los archivos DJVU.

¹⁴ Este programa sólo tiene interfaz en el idioma ruso.

¹⁵ Es un convencionalismo ruso el que la numeración de las páginas comience desde la primera página del libro. En las artes gráficas occidentales, se acostumbra señalar por separado las páginas preliminares y con números romanos, por lo tanto la compensación no será de 1 sino estará entre 10 y 20.

De igual forma se crean los hiperenlaces en el índice de materias. Es necesario seleccionar una entrada diferente a la predeterminada en la lista desplegable del campo . La entrada predeterminada que se ve en la lista significa “Índice general”. Las otras entradas quieren decir que desea procesar el índice de materias. Se aplican las mismas configuraciones.

Después de terminar el proceso, vea el archivo DJVU y compruebe que los hiperenlaces se agregaron correctamente. El programa depende del texto OCR para determinar los números de página de los hiperenlaces. Así que cualquier error en el OCR lleva el hiperenlace a posiciones u objetivos erróneos.

A. ¿Dónde descargar los programas?

<i>Nombre del programa</i>	<i>Sitio de descarga</i>	<i>Condición</i>
IrfanView 4.1	www.irfanview.com	Gratis
ScanKromsator 5.9	www.djvu-soft.narod.ru	Gratis
DjvuSolo 3.1	www.djvu-soft.narod.ru	Gratis
Djvu Editor 4.x, 5.x, 6.x (DDE/DEE)	www.djvu-soft.narod.ru	Pago
FineReader 7.x, 8.x	www.abbyy.com	Prueba
DjvuOCR 2.2 beta	djvuocr.ucoz.ru	Gratis
Djvu Hyperlinks Editor	www.djvu-soft.narod.ru	Gratis

¡Muchas gracias **monday2000** por la creación del sitio djvu-soft.narod.ru!

Nota para los usuarios de Linux: todos los programas de esta tabla trabajan razonablemente bien en el emulador de Windows estándar (wine). Sin embargo, algunos programas (IrfanView, DDE/DEE, FineReader) pueden fallar al instalarse si se ejecutan sus "setup.exe" respectivos. Deberá conseguir las versiones portátiles ("portable" o "installed") de dichos programas, ya que esas no necesitarán ejecutar ningún instalador.

Índice

calidad, 2

DJVU, 3, 18

 diccionario, 19

 capa OCR, 21

 reorganizar páginas, 24

enderezado, 9

escaneo, 9, 10

 espacio en disco, 9

 escala de grises, 2

 con cámara digital, 5

FineReader

 problemas, 3

ilustraciones, 2

ilustraciones en color, 23

instantáneas en ruso, 1

IrfanView, 6

JPEG, 6

 alteraciones digitales, 7

 problemas con textos, 7

kromsating, 11

ScanKromsator, 3, 10

 cortadores, 12

 ejecución en borrador, 11

 ejecución definitiva, 16

 zona de ilustraciones, 17

sobremuestreo, 2, 16

TIFF, 6

uso de Linux, 27