

ReadirisTMPRO9

MANUAL DEL USUARIO



Readiris Pro

© 2004 I.R.I.S. Reservados todos los derechos
Tecnología OCR propiedad de I.R.I.S.

Tecnología Connectionist, AutoFormat y tecnología lingüística propiedad de I.R.I.S.

© 2004 I.R.I.S. Reservados todos los derechos

¡AHORRE TIEMPO, NO VUELVA A TECLEAR!

Queremos darle la enhorabuena por haber adquirido Readiris. Este *software* le resultará, sin duda, de gran ayuda a la hora de capturar textos, tablas y gráficos.

Por eficientes que sean los ordenadores, primero tiene que teclear su información. Si alguna vez ha tenido que volver a teclear un informe de 15 páginas o una amplia tabla de cifras, sabrá lo tedioso y prolongado que puede ser. Utilice este paquete de *software* OCR de punta para introducir automáticamente el texto en sus aplicaciones y adquirirá un nivel sin precedentes de eficiencia y comodidad.

Escanee un documento impreso o mecanografiado, indique las zonas de interés - o deje que el sistema las detecte - y ejecute el reconocimiento de caracteres. Los documentos de múltiples páginas se procesan de principio a fin de una sola vez. Unos cuantos clics con el ratón equivalen a largas horas de trabajo, puesto que Readiris convierte sus documentos en papel en archivos digitales modificables hasta 40 veces más rápido que si volviese a teclearlos.

El asistente de OCR lo ayudará durante el proceso de reconocimiento: ¡responda a unas simples preguntas para obtener rápidos resultados con Readiris! Podrá enviar directamente el resultado del reconocimiento a un procesador de textos o a una hoja de cálculo. Para reconocer un fax y convertir documentos PDF, basta con arrastrar el archivo desde el Explorador de Windows hasta la ventana de la aplicación Readiris. O hacer clic sobre cualquier imagen con el botón secundario del ratón para enviarla inmediatamente a Readiris.

Readiris reconoce datos tabulados y los reproduce como hojas de cálculo o como objetos de tabla en un procesador de textos; los datos numéricos están así listos para ser procesados.

Basado en la tecnología Connectionist de I.R.I.S., Readiris representa lo mejor que puede ofrecer el reconocimiento de caracteres. La extracción de características independiente de las fuentes está complementada por técnicas de autoaprendizaje derivadas de una red neural patentada. El sistema puede aprender nuevos caracteres mediante análisis del contexto: los conocimientos lingüísticos sobre sílabas y palabras mejoran el rendimiento del reconocimiento.



Readiris reconoce hasta 107 idiomas: todos los idiomas americanos y europeos, incluyendo los idiomas centroeuropeos, bálticos, cirílicos (“rusos”) y el griego. (Opcionalmente, puede leer cuatro idiomas asiáticos - japonés, chino simplificado y tradicional y coreano.) Readiris puede procesar incluso alfabetos mixtos: el programa detecta las palabras “occidentales” que aparecen en los documentos griegos, cirílicos o de lenguas asiáticas, ya que muchos nombres propios, marcas, etc. no se transcriben y aparecen en caracteres latinos.

Readiris usa la lingüística *durante* la fase de reconocimiento, no después. Como resultado de ello, Readiris reconoce documentos de todo tipo con máxima precisión, incluyendo documentos de baja calidad, faxes e impresiones con matriz de agujas. Procesa los documentos mal escaneados y copiados que contienen formas de fuentes demasiado claras u demasiado oscuras. Los caracteres unidos entre sí (“ligaduras”) se resuelven y las formas fragmentadas, como los símbolos de las matrices de agujas, se vuelven a componer.

La ventana de verificación por parte del usuario no sólo destaca los caracteres dudosos, sino que también aumenta la precisión del sistema. Todas las soluciones confirmadas por el usuario se almacenan en memoria, aumentando así la velocidad y confianza a medida que trabaja. Usar Readiris significa hacerlo más inteligente cada vez! Esta potente herramienta de aprendizaje le permite enseñar a Readiris para que reconozca caracteres especiales, como símbolos matemáticos y dingbats, e incluso procesar fuentes distorsionadas como las que se encuentran en los documentos de la vida real.

Para aumentar aún más su productividad, Readiris no sólo reconoce textos, sino que también puede darles *formato*. Si utiliza el “formato automático”, Readiris reproducirá una copia facsímil del documento escaneado, manteniendo el formato de las palabras, párrafos y páginas del documento original.

Se utilizan tipos de letras similares, los tamaños y estilos usados en el documento original se mantienen tras el reconocimiento. La colocación de columnas, bloques de texto y gráficos imita la de los documentos originales. Y como Readiris escanea sin ningún problema imágenes en escala de grises o en color, puede capturar cualquier gráfico - ya sean dibujos, fotografías en blanco y negro o ilustraciones

en color. Cuando el documento contiene tablas, Readiris las recoloca imitando perfectamente las celdas y reproduciendo los bordes de las tablas originales.

En otras palabras, Readiris le permite archivar una copia fiel de sus documentos en forma de archivos digitales de texto que puede modificar en lugar de una imagen escaneada. Existen varios niveles de formato, la elección la tiene el usuario.

Readiris permite utilizar una amplia gama de escáneres: escáneres planos, escáneres con alimentador de hojas, periféricos multifuncionales (“*MFP*”) y cámaras digitales. Readiris también cumple la norma Twain y, además, puede utilizarse en algunas plataformas de escaneo.

TABLA DE MATERIAS

¡Ahorre tiempo, no vuelva a teclear!	III
Tabla de materias	V
Créditos y derechos de autor	VII

Capítulo 1: Instalación

Requisitos del sistema	1-1
Instalación del software de Readiris	1-1
Desinstalación del software de Readiris	1-4
Programa de desinstalación de Readiris.....	1-5
Programa de (des)instalación de Windows	1-5
Instalación de software opcional	1-6
Instalación de productos conexos	1-8
Archivos instalados	1-10
Archivo “Léame” y documentación	1-10
Controladores de escáner	1-10
¡Regístrese!	1-10
Cómo obtener asistencia técnica	1-12

Capítulo 2: Visita guiada

Inicio del software	2-1
Primera ejecución	2-2
Descubrir el interfaz de Readiris	2-3



Primera lección introductoria	2-6
Ampliación de imágenes	2-10
Uno, descomponer una imagen escaneada	2-13
Una y media, clasificación de ventanas	2-15
Dos, trazar ventanas manualmente en una imagen	2-18
Tres, guardar modelos de ventanas	2-23
Readiris lo lleva por todo el mundo	2-24
Readiris cambia de idioma cuando es necesario	2-30
Definición de las características del documento	2-33
¡Readiris se hace cada vez más inteligente!	2-35
Aprender	2-38
Descartar Aprendizaje	2-38
Borrar	2-39
Deshacer	2-39
Terminar	2-39
Cancelar	2-39
Papel de los diccionarios de fuentes	2-40
Envío de resultados a la aplicación de trabajo	2-43
Guardar los resultados en un archivo de texto	2-47
Creación de documentos transportables	2-51
... o reconocimiento de dichos documentos	2-56
Reconocimiento de varias páginas	2-57
Edición de documentos con múltiples páginas	2-65
Abrir un nuevo documento	2-67
Reconocimiento de zonas de texto	2-68
Organización del texto resultante	2-70
Configuración del escáner	2-71
Ponga colores en sus textos digitalizados!	2-73
Aparatos distintos, resolución distinta	2-76
Cómo guardar las propiedades predeterminadas	2-80
Cómo guardar propiedades específicas	2-81
Numerizar documentos	2-82
Ajuste de las imágenes escaneadas	2-85
Deje que el asistente trabaje por usted	2-90
Readiris reproduce el diseño de sus documentos	2-91
Utilice columnas en lugar de cuadros	2-96
Formato de textos, 2ª parte	2-98

Exportar el texto varias veces	2-99
Guardar gráficos por separado	2-100
Lectura de faxes y reconocimiento diferido	2-102
Reconocimiento de tablas	2-104
Cómo obtener ayuda en línea	2-110

CRÉDITOS Y DERECHOS DE AUTOR

El programa Readiris ha sido diseñado y desarrollado por I.R.I.S. La tecnología OCR, Connectionist, AutoFormat y tecnología lingüística son propiedad de I.R.I.S. Asimismo, I.R.I.S. detenta los derechos de autor del *software* Readiris, de la tecnología OCR, de la tecnología lingüística, del sistema de ayuda en línea y del presente manual.

AutoFormat, Cardiris, Connectionist, el logotipo de I.R.I.S., Readiris y la tecnología lingüística I.R.I.S. son marcas de I.R.I.S.

Analizador sintáctico XML desarrollado por Apache. Este producto incluye software desarrollado por la Apache Software Foundation (www.apache.org).

Acrobat y Reader son marcas (registradas) de Adobe. AsianBridge es una marca de TwinBridge. AsianSuite es una marca de UnionWay. Excel, Windows y Word son marcas de Microsoft. Intel es una marca registrada de Intel.



Capítulo 1

INSTALACIÓN

Este capítulo describe los requisitos del sistema y la instalación del *software* Readiris.

REQUISITOS DEL SISTEMA

Ésta es la configuración mínima del sistema necesaria para utilizar Readiris:

- ☐ un procesador Intel 486 o compatible. Se recomienda un procesador Pentium.
- ☐ 32 MB de RAM. Se recomiendan 64 MB para procesar imágenes en escala de grises o en color.
- ☐ 110 MB de espacio en el disco duro. Bastarán 95 MB si deja los archivos de muestra en el CD-ROM.
- ☐ sistema operativo Windows XP, Windows ME, Windows 2000, Windows 98 o Windows NT 4.0.

Readiris Corporate requiere un monitor con una resolución de 1024 x 768.

Puede que algunos **controladores de escáner** no funcionen con la(s) última(s) versión(es) de Windows. Consulte la documentación del escáner para ver qué plataformas pueden utilizarse.

INSTALACIÓN DEL SOFTWARE DE READIRIS

El programa Readiris se entrega exclusivamente en un **CD-ROM autoejecutante**. Para instalarlo, simplemente introduzca el CD-ROM en la unidad

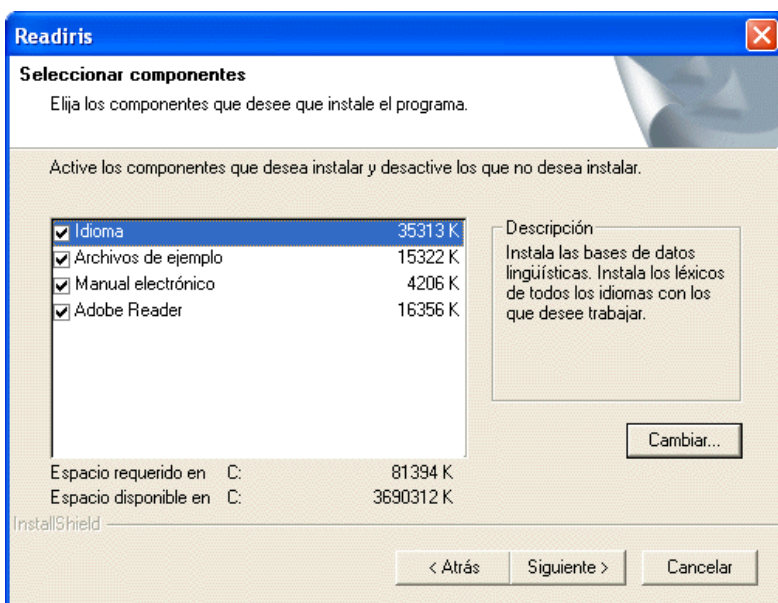


de CD-ROM y espere a que el programa de instalación comience a ejecutarse. Siga las instrucciones que aparecen en la pantalla.

Si la instalación no comienza a ejecutarse con el CD-ROM introducido en la unidad de CD-ROM, ejecute el programa MENU.EXE para instalar el programa.

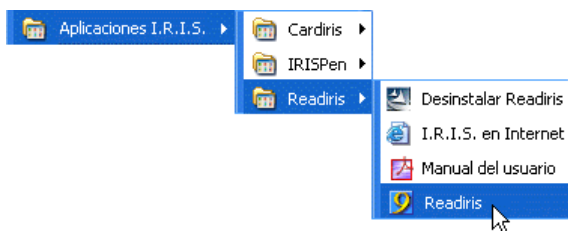
Los usuarios de Windows XP, Windows 2000 y Windows NT deben comprobar que tienen los **derechos de acceso** adecuados. Si fuera necesario, deberán ponerse en contacto con el administrador del sistema.

En el momento de la instalación se proponen algunas opciones. Asegúrese de instalar las **bases de datos lingüísticas** de todos los idiomas con los que piensa trabajar. Por defecto, se instalan todas. Le recomendamos que también instale las **imágenes de muestra** que se usan en los tutoriales del presente manual.



De igual manera, asegúrese de instalar Adobe Reader para acceder a la documentación del programa Readiris en caso necesario. El **manual electrónico** se copia por defecto en el disco duro, pero puede dejarlo en el CD-ROM.

El programa de instalación crea automáticamente el submenú "Aplicaciones I.R.I.S. - Readiris" en el menú "Programas".



Hará lo mismo creando un acceso directo a Readiris en el escritorio de Windows. Así podrá usted lanzar Readiris directamente desde su escritorio.

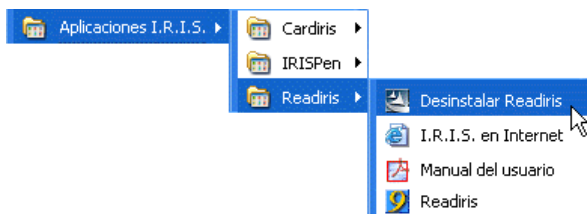


DESINSTALACIÓN DEL SOFTWARE DE READIRIS

Sólo existen dos formas correctas de desinstalar Readiris: utilizando el programa de desinstalación de Readiris y utilizando el programa automático de (des)instalación de Windows. Le recomendamos encarecidamente *no* desinstalar Readiris o sus módulos de *software* borrando manualmente los archivos de programa.

Programa de desinstalación de Readiris

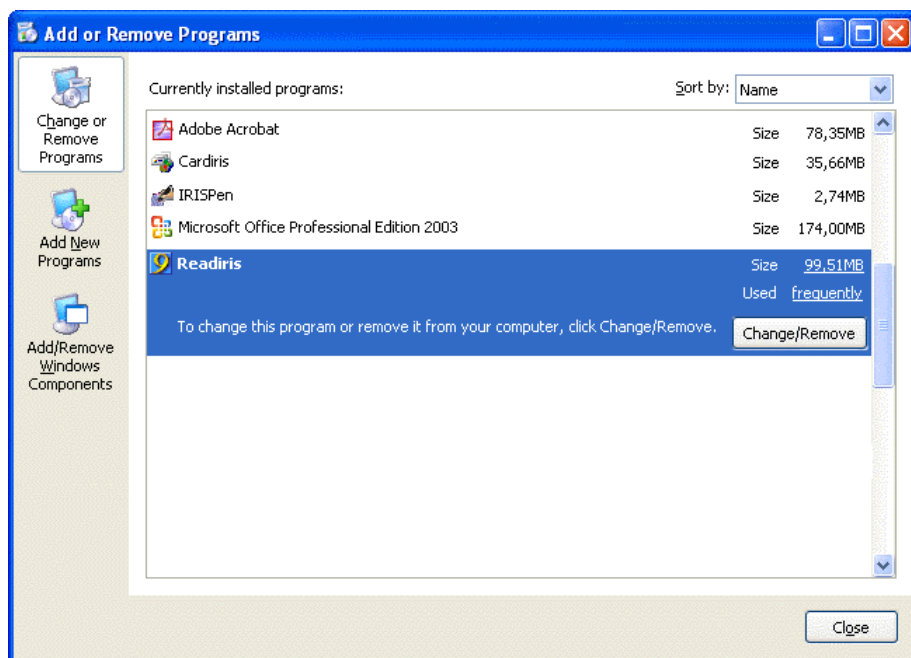
Seleccione "Desinstalar Readiris" en el submenú "Aplicaciones I.R.I.S. - Readiris" para ejecutar el programa de desinstalación de Readiris y siga las instrucciones que aparecen en pantalla.



Programa de (des)instalación de Windows

Ejecute los siguientes pasos para utilizar el programa automático de (des)instalación de Windows.

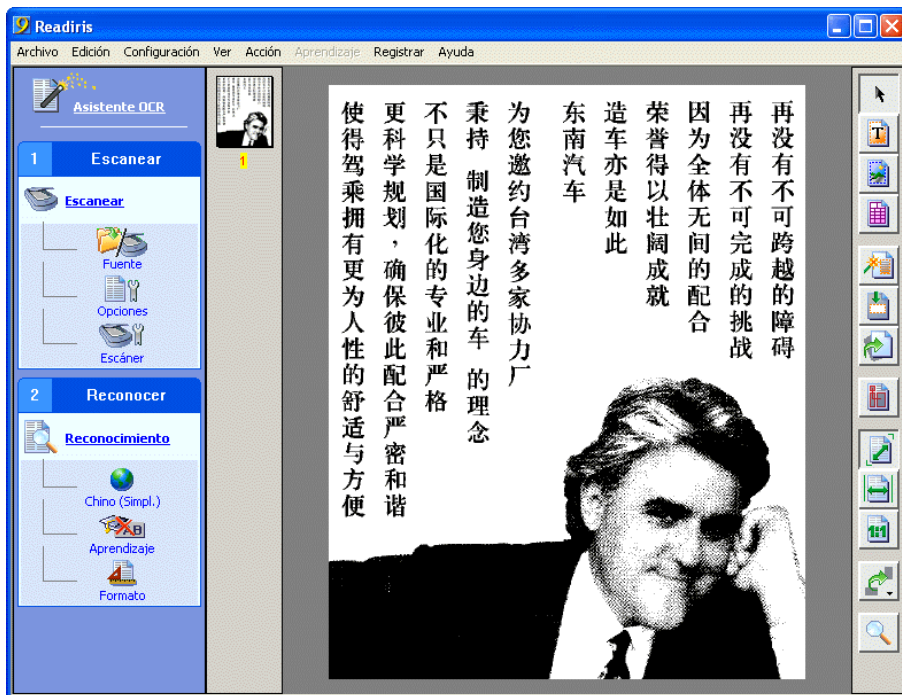
- ☐ Haga clic en "Configuración" en el menú "Inicio" de Windows y vaya al "Panel de Control".
- ☐ Haga clic en el "Agregar o Quitar Programas" situado en el panel de control.



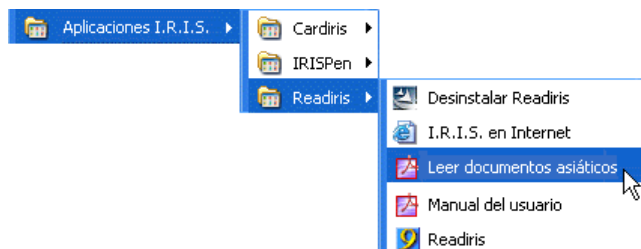
- ☐ Siga las instrucciones que aparecen en pantalla para eliminar el *software* de Readiris.

INSTALACIÓN DE SOFTWARE OPCIONAL

Existe un *software* opcional complementario a Readiris. Se trata del módulo OCR para lenguas asiáticas, que le permitirá reconocer documentos en japonés, chino simplificado, tradicional y coreano. Este *software* se suministra asimismo en un CD-ROM autoejecutable.



Al instalar esta opción, se habilitará la documentación específica en la que se explica cómo reconocer documentos escritos en lenguas asiáticas.

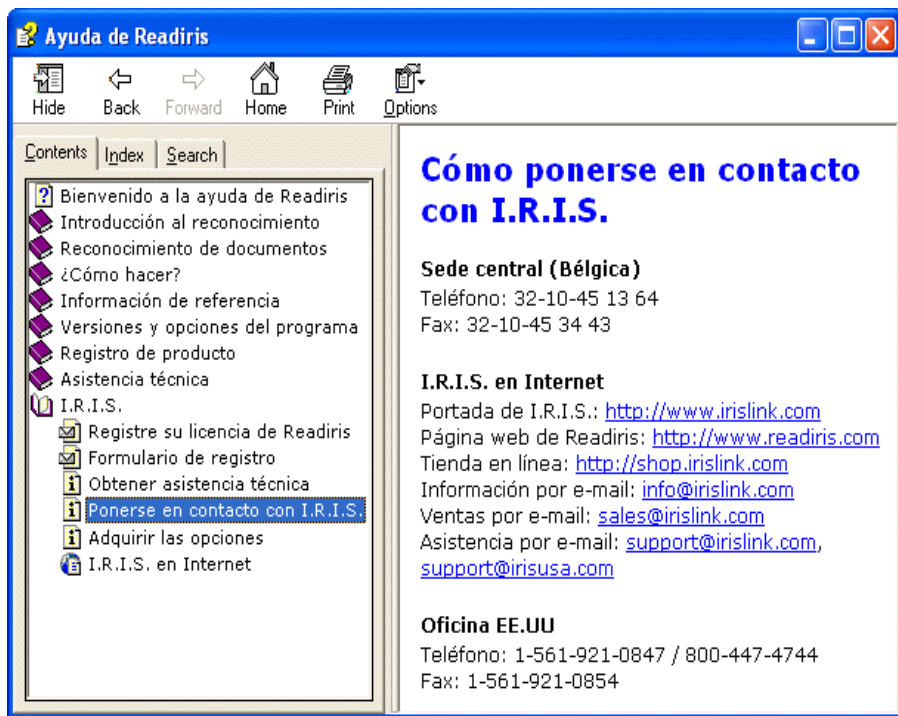


INSTALACIÓN DE PRODUCTOS CONEXOS

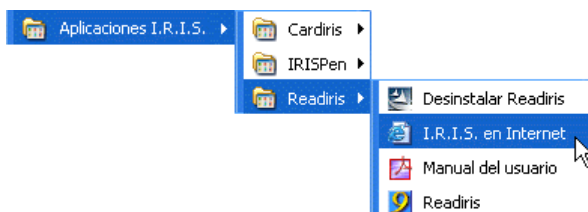
Según del paquete de *software* que haya adquirido, Readiris puede suministrarse con una versión de evaluación del producto Cardiris, un **organizador de tarjetas de visita**.

Si su versión de Readiris contiene este programa gratuito, también se instalará mediante el programa autoejecutable del CD-ROM. Siga para ello las instrucciones que aparecen en pantalla.

Póngase en contacto con I.R.I.S. para obtener más información acerca de este *software*, la opción "Contactar con I.R.I.S." del menú "Ayuda" de Readiris describe las formas en que puede ponerse en contacto con I.R.I.S.



Un icono del submenú "Aplicaciones I.R.I.S. - Readiris" del menú "Programas" lo conducirá directamente a la **página web** de I.R.I.S. También podrá acceder a esta página desde la pantalla de inicio del programa, o seleccionando la opción "I.R.I.S. en Internet" del menú "Ayuda".



ARCHIVOS INSTALADOS

El programa de instalación crea una carpeta en la que se localizan los archivos de Readiris. Nunca intente desinstalar Readiris o alguno de sus módulos borrando manualmente los archivos de programa, utilice el programa de desinstalación de Readiris o el programa automático de (des)instalación de Windows. Véase más arriba.

Archivo “Léame” y documentación

README.HTM Archivo “Léame” (en formato HTML)

MANUAL.PDF Manual del usuario (en formato Adobe Acrobat)

Controladores de escáner

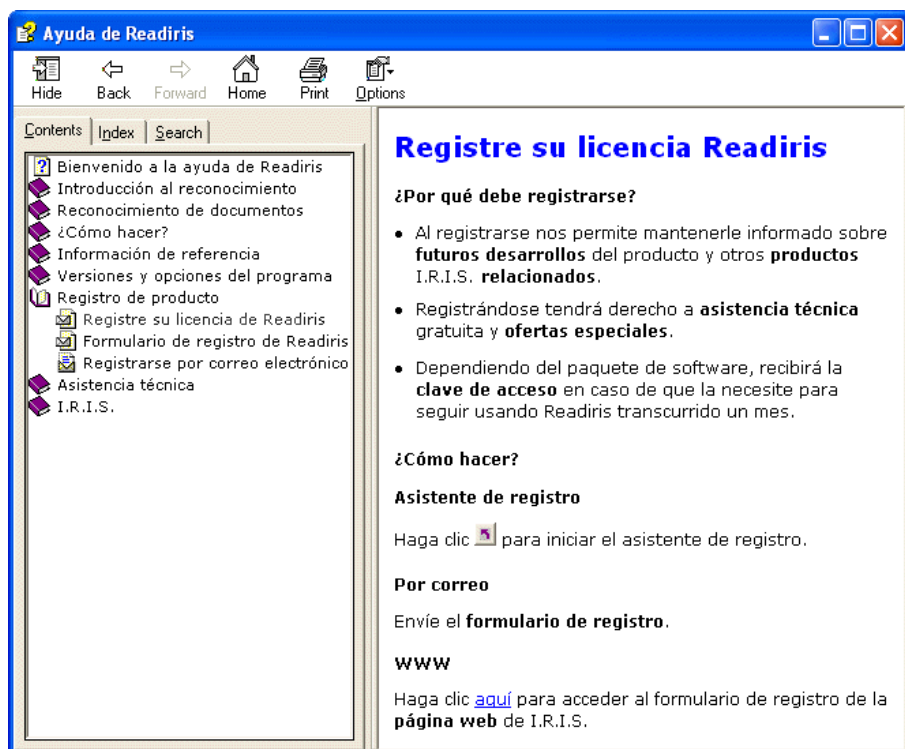
No dude en ponerse en contacto con el fabricante del equipo o con su distribuidor si siguiera teniendo problemas con los controladores del escáner. La mayoría de fabricantes permiten descargar las últimas versiones de los controladores directamente desde su página web.

¡REGÍSTRESE!

No olvide de registrar su licencia Readiris. Si lo hace lo mantendremos informado sobre futuros productos y productos conexos de I.R.I.S. Las ventajas

de registrarse, como la **asistencia gratuita** y las **ofertas especiales**, están reservadas estrictamente a los usuarios registrados.

Puede registrarse de muchas formas: enviándonos por correo el formulario de registro, o por fax el formulario electrónico, llamando a I.R.I.S. en horas de oficina, o bien completando el formulario de registro en la página web de I.R.I.S.



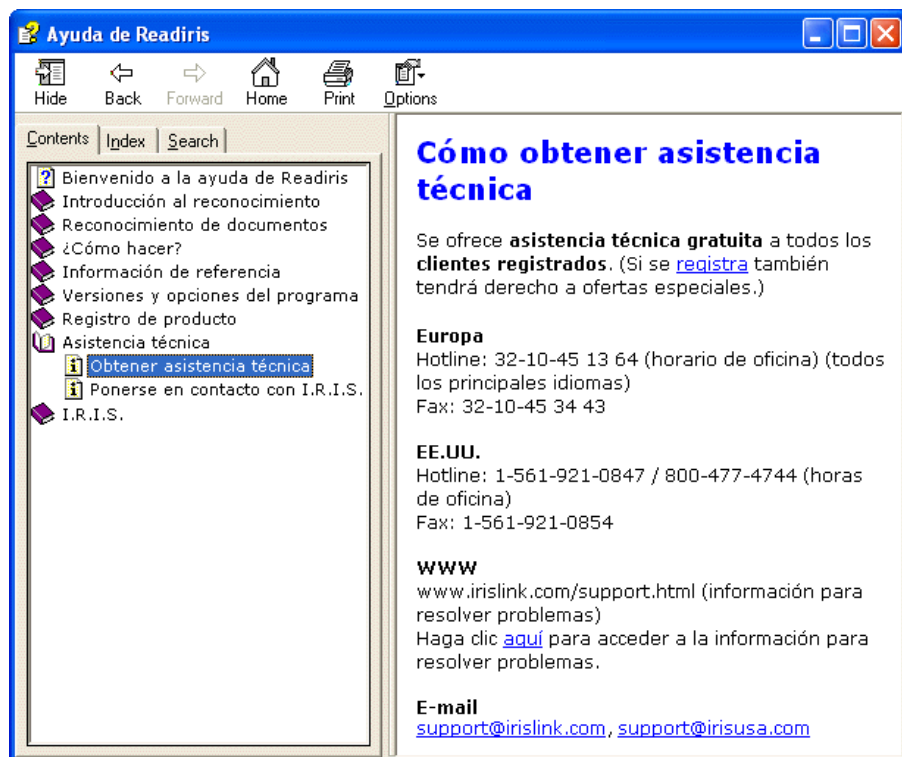
El **asistente de registro** de Readiris, que encontrará en el menú "Registrar" del programa Readiris, lo guiará cómodamente durante todo el proceso.



Dependiendo de la versión adquirida, recibirá la **clave de acceso** necesario para continuar utilizando el programa Readiris después de un mes de demostración.

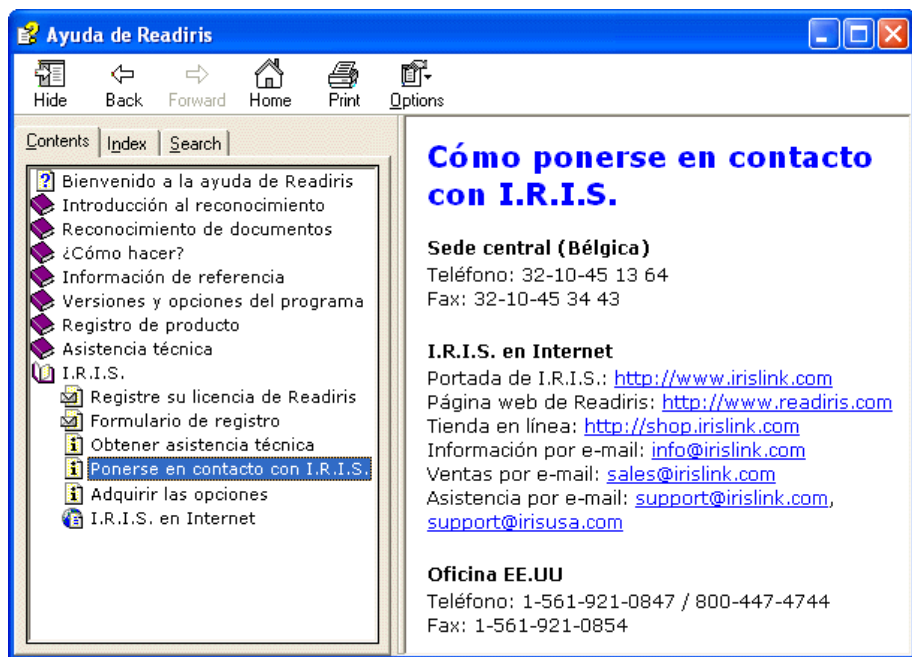
CÓMO OBTENER ASISTENCIA TÉCNICA

La opción "Asistencia Técnica" del menú "Ayuda" de Readiris explica cómo obtener asistencia técnica. Describa claramente lo que le sucede e incluya toda la información pertinente sobre Readiris, su escáner y su sistema informático.

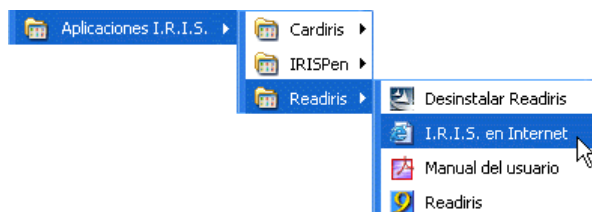


PÓNGASE EN CONTACTO CON I.R.I.S.

También puede ponerse en contacto con I.R.I.S. para obtener más información sobre otras soluciones de su gama de productos; el mando "Contactar con I.R.I.S." del menú "Ayuda" de Readiris describe las formas en que puede ponerse en contacto con I.R.I.S.



Un icono del submenú "Aplicaciones I.R.I.S. - Readiris" del menú "Programas" lo conducirá directamente a la **página web** de I.R.I.S. También podrá acceder a esta página desde la pantalla de inicio del programa, o seleccionando el mando "I.R.I.S. en Internet" del menú "Ayuda".





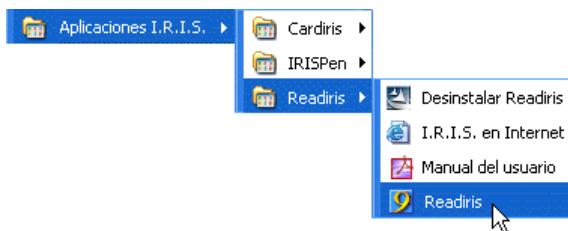
Capítulo 2

VISITA GUIADA

Readiris es un paquete con la tecnología OCR de punta provisto de numerosas características avanzadas. En este capítulo describiremos todas sus principales características y añadiremos consejos sobre el uso de Readiris.

INICIO DEL SOFTWARE

Haga clic en el submenú "Aplicaciones I.R.I.S. - Readiris", o en el acceso directo a la aplicación Readiris situada en su escritorio.



Se visualizarán la pantalla de inicio y la ventana de la aplicación Readiris. La pantalla de inicio muestra la versión y los derechos de autor del programa Readiris. Le permite también un acceso directo a la **portada web** de I.R.I.S. - haga clic en la dirección URL para acceder a la página. Esta pantalla desaparecerá haciendo clic en cualquier otra zona.



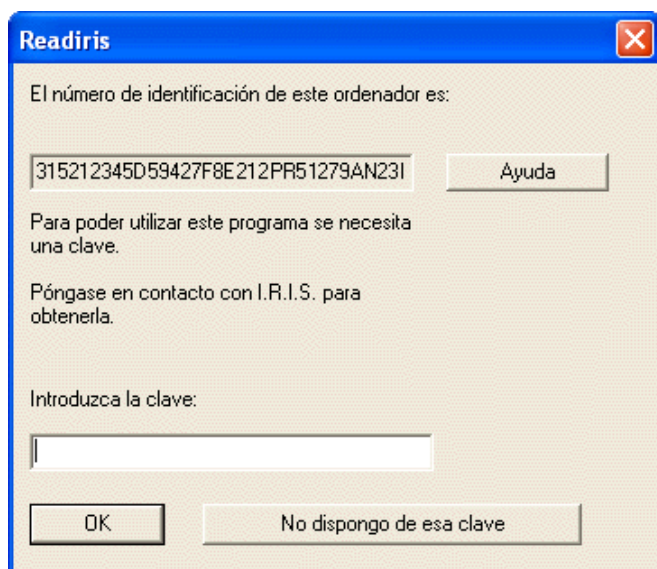
La siguiente ventana se refiere al asistente de OCR. Haga clic en "Cancelar".

PRIMERA EJECUCIÓN

Dependiendo del paquete de *software* que haya adquirido, la primera ejecución puede ser especial, pues se le pedirá que registre su licencia.

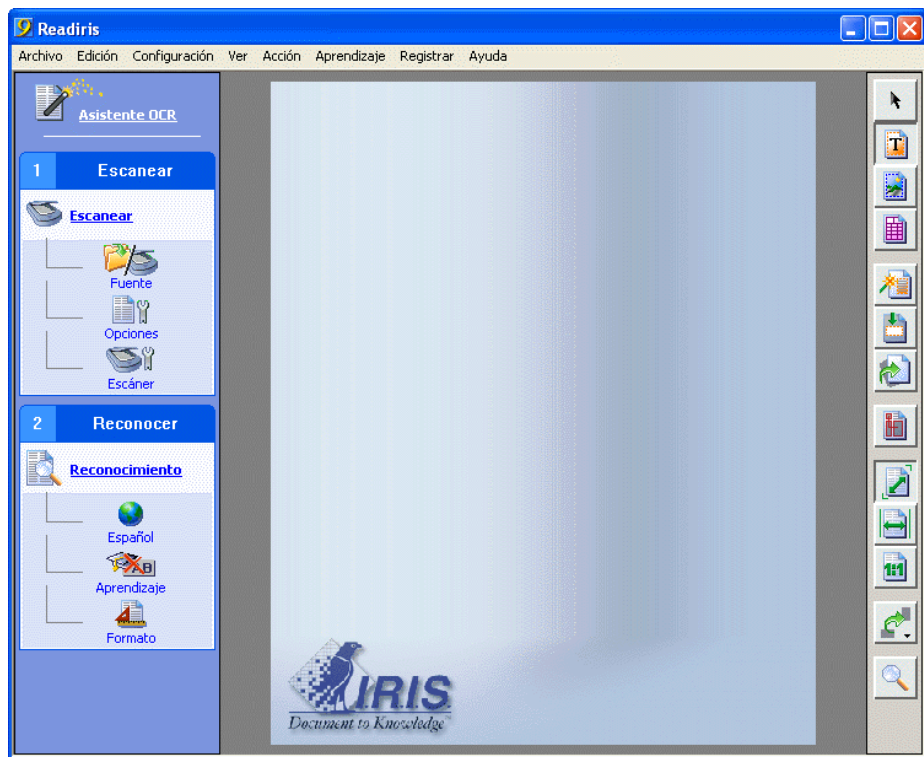
En tal caso, la utilización de Readiris está limitada a un mes y, al registrarse, recibirá de I.R.I.S. una **clave de acceso** gratuita para continuar utilizando el *software* transcurrido dicho plazo.

Necesitará su **número de identificación** para generar la clave de acceso; asegúrese de disponer de dicho número o de mencionarlo al registrar su licencia.



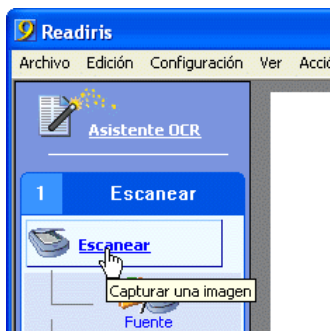
DESCUBRIR EL INTERFAZ DE READIRIS

La ventana de aplicación de Readiris no sólo contiene **menús de opciones**, sino también dos barras de botones que permiten acceder rápidamente a todas las opciones frecuentes. Al principio, algunos menús están difuminados, pues se refieren a la vista preliminar. Mientras no se abra una imagen, no se puede disponer de ellos.



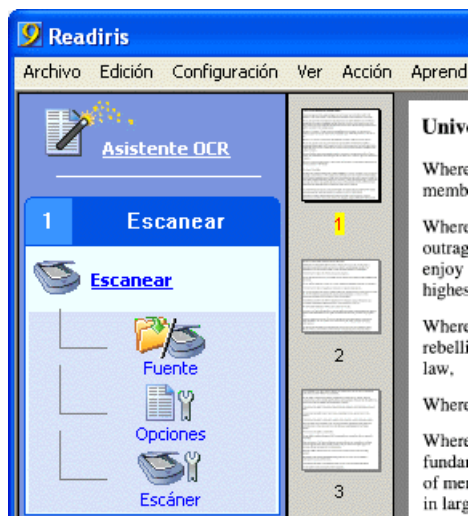
Otro tanto sucede con la **barra de herramientas de imágenes** situada a la derecha de la ventana de aplicación: contiene todas las opciones que necesita durante la vista preliminar de la imagen. La **barra principal de herramientas** situada a la izquierda permite acceder rápidamente a todas las opciones generales usadas con frecuencia.

Para saber la opción que corresponde a cada botón, coloque el cursor del ratón sobre este último durante un momento: una **etiqueta emergente** le dirá lo que hace el botón.



La parte interior de la ventana o **zona de imagen** es donde se muestran las imágenes escaneadas. Puede arrastrar los archivos gráficos hacia la zona de imágenes (o hacia el icono de Readiris) para reconocerlos.

Cuando se procesan páginas, en la parte izquierda aparecerá una nueva barra de herramientas, denominada **barra de herramientas de páginas**. En esta barra se representan las distintas páginas del documento y permite el acceso a las funciones de página haciendo clic con el botón secundario del ratón (se abre el menú Contexto).



PRIMERA LECCIÓN INTRODUCTORIA

La mejor manera de familiarizarse con el manejo de Readiris es sin duda utilizándolo. El programa contiene una serie de **imágenes ya escaneadas** que le permitirán comenzar incluso si no tiene un escáner conectado a su ordenador. Pasemos a estas imágenes.

El botón "Fuente" de la barra principal de herramientas determina si va a utilizar un escáner o una imagen ya escaneada como origen de la imagen.

Readiris reconoce igualmente imágenes en color, en escala de grises y en blanco y negro. Readiris le permite abrir documentos PDF Adobe Acrobat y archivos de imágenes JPEG, Paintbrush (PCX), fax DCX (una versión con múltiples páginas del formato Paintbrush), PNG, TIFF (no comprimidos, comprimidos en LZW, packbits, Grupo 3 y Grupo 4), TIFF con múltiples páginas y mapas de bits de Windows (BMP).

Esta función resulta especialmente útil para convertir sus **faxes** en archivos digitales de texto.

Cuando vaya a abrir una imagen ya escaneada, con el botón "Fuente" de la barra principal de herramientas seleccione el disco y no el escáner como origen de la imagen.



A continuación haga clic en el botón "Abrir". (Cuando selecciona el disco como origen de la imagen, al botón "Escanear" lo sustituye el botón "Abrir". Asimismo, el mando "Escanear" del menú "Acción" la sustituye el mando "Abrir".)



A continuación haga clic en el botón "Abrir" del menú "Archivo" y abre una imagen escaneada previamente - esto es posible incluso si su escáner funciona como origen actual de la imagen.



Se le pedirá que elija un archivo gráfico. Seleccione el archivo SPANISH.JPG de la carpeta Readiris. Como se trata de una imagen en color, además de leerla directamente desde el disco, se crea una versión "binaria", en blanco y negro, para el proceso de OCR.

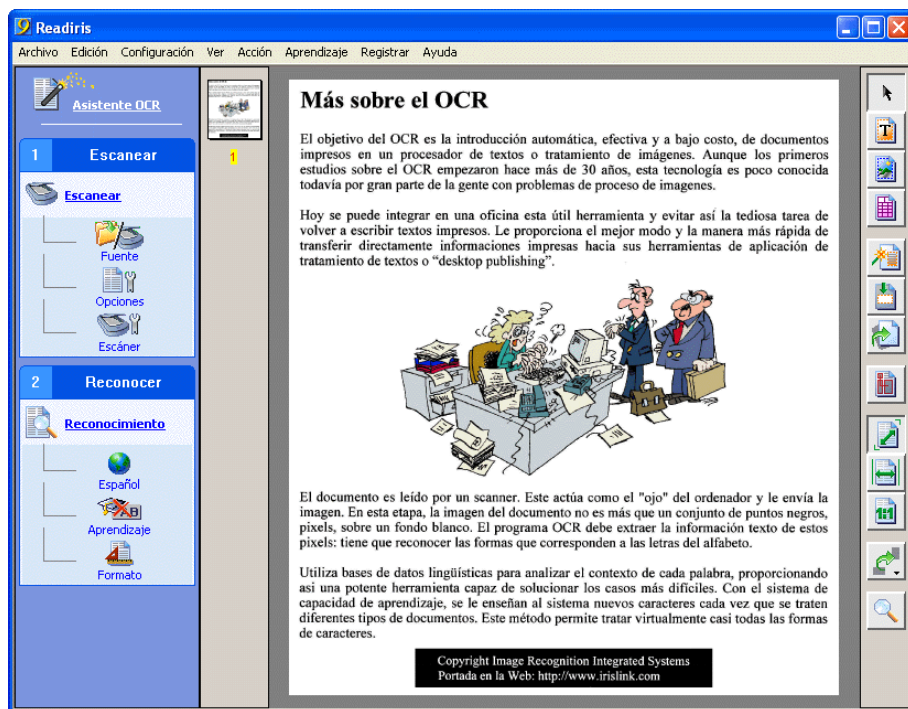


Cargando D:\Readiris\spanish.jpg

Convirtiendo

44%

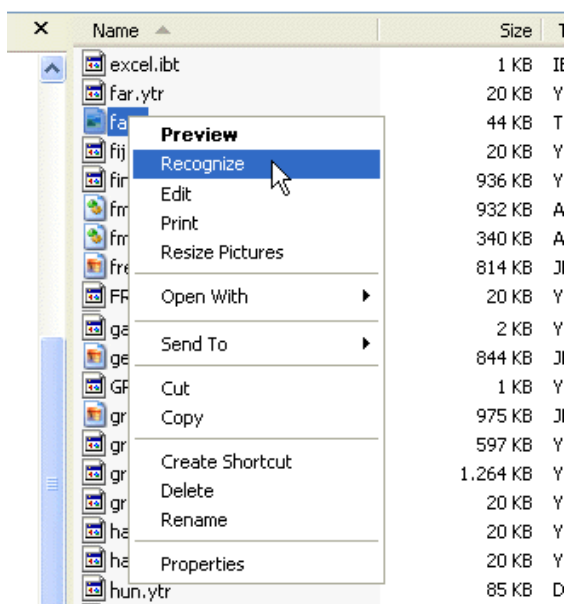
Finalmente, la imagen se visualiza en la zona de imágenes y la barra de herramientas de páginas indica que una página está activa en Readiris.



Una tercera forma de abrir imágenes ya escaneadas es el “**arrastre con el ratón**”: arrastre imágenes desde el Explorador de Windows hacia la zona de imagen de Readiris o hacia el icono de Readiris y se abrirán inmediatamente.



Por último, podrá abrir imágenes directamente desde el Explorador de Windows **haciendo clic con el botón derecho** del ratón en un archivo gráfico y seleccionando el mando "Reconocer" del menú Contexto. (Este mando sólo estará disponible si el archivo en cuestión pertenece a alguno de los reconocidos por el programa.)



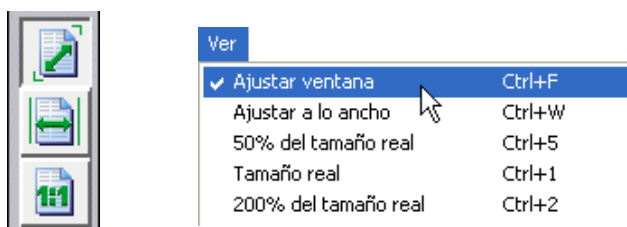
Esto no significa que el OCR se efectúe inmediatamente. Con el fin de proporcionar al usuario la máxima flexibilidad de uso, únicamente se ejecuta la aplicación Readiris y se abre la imagen.

La barra de herramientas de imágenes situada a la derecha de la ventana de aplicación de Readiris contiene todas las opciones que necesita durante la vista preliminar de la imagen: herramientas para indicar las zonas de interés, para girar la imagen, para ampliarla o reducirla, etc.

AMPLIACIÓN DE IMÁGENES

Readiris cuenta con varias opciones que le permiten **ampliar** la imagen escaneada, por ejemplo, para verificar la calidad de escaneado.

La barra de herramientas de imágenes contiene botones que le permiten ampliar a tamaño real, ajustar la imagen al ancho de página y ajustar toda la imagen en la ventana de vista preliminar. El menú "Ver" contiene las mismas opciones y añade dos niveles adicionales de ampliación: puede mostrar la imagen al 50% y al 200% de su tamaño real. En tamaño real, un pixel de pantalla corresponde a un pixel de la imagen. (Existen macros para todos los niveles de ampliación.)

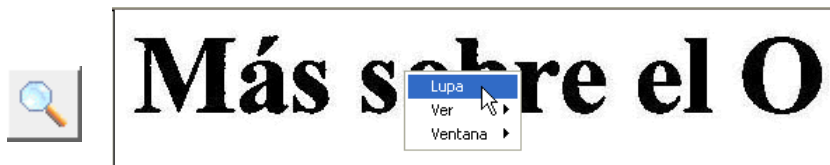


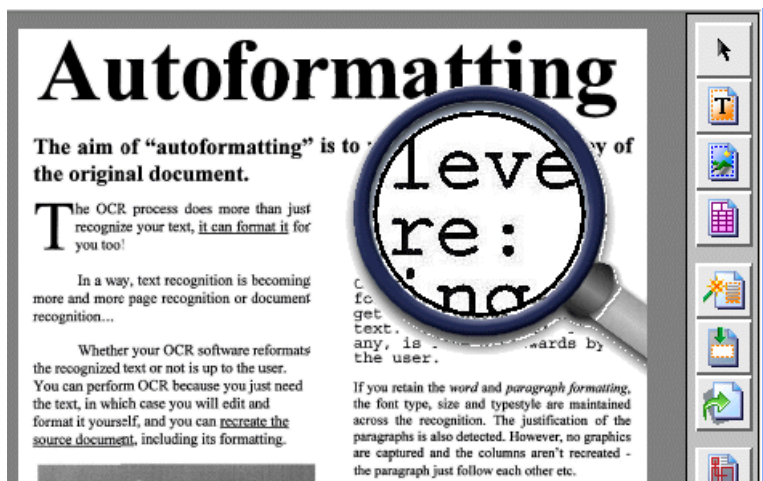
Observe también que dispone de un acceso a los niveles de ampliación gracias al clic derecho. Haga clic con el botón derecho del ratón para que aparezca el menu contextual y seleccione el nivel de ampliación adecuado.



Además, puede hacer *doble* clic con el botón derecho del ratón en una zona de la imagen escaneada para ampliarla inmediatamente a tamaño real. Repita la operación para volver a disminuirla.

Por último, puede utilizar la **lupa** para ampliar detalles del documento escaneado. La lupa de aumento también está disponible en el menú contextual al hacer clic sobre la imagen con el botón derecho del ratón.

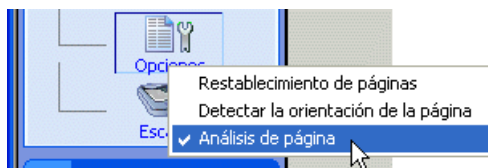




UNO, DESCOMPONER UNA IMAGEN ESCANEADA

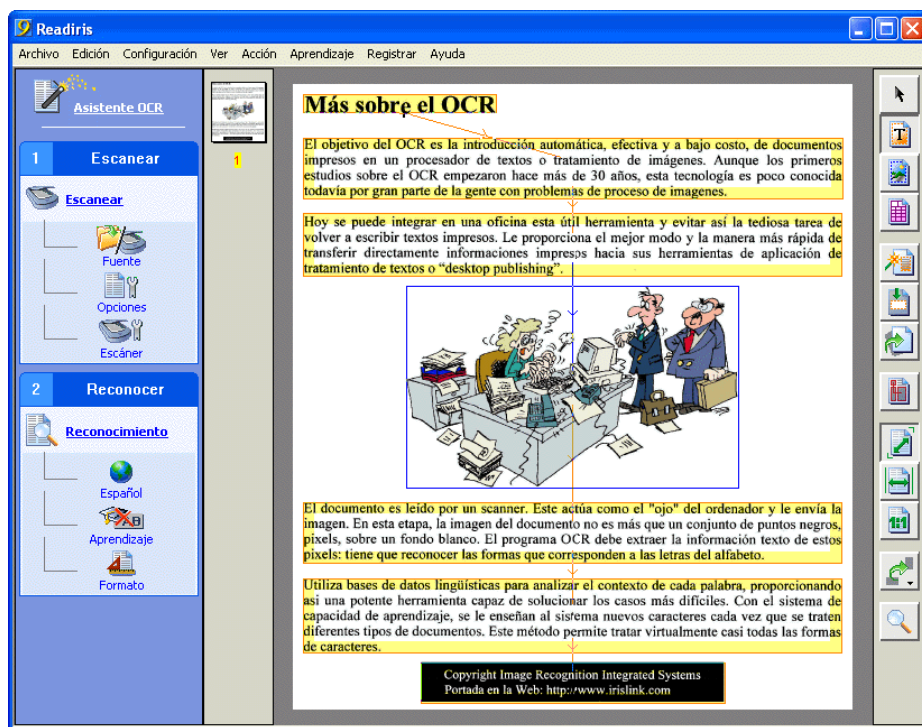
Una vez escaneada la imagen debe indicar las partes que desea convertir en texto modificable trazando marcos, denominados "ventanas", alrededor de las zonas de interés.

De hecho, Readiris lo hará automáticamente si activa la opción "Análisis de Página" desde el botón "Opciones" de la barra de herramientas principal (o desde el menú "Configuración").





La descomposición automática de páginas resulta especialmente útil para reconocer **textos con columnas** y documentos con un diseño de página complejo (incluirán posiblemente gráficos y tablas).



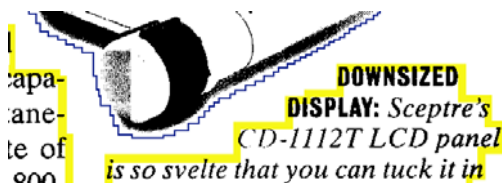
La división de página se realiza mediante tres **tipos de ventanas**: ventanas de texto, de gráficos y de tablas. Readiris separa en cada página bloques de texto, zonas de tablas y zonas de gráficos (fotografías, ilustraciones, etc.). (La manera de reconocer tablas y guardar gráficos se tratará ampliamente más adelante.)

Un **código de color** indica el tipo de ventana de que se trata: las zonas de texto tienen un borde amarillo, las ventanas gráficas tienen un borde azul y las de tablas, un borde morado.

El número de ventanas se indica en todo momento en las etiquetas emergentes de las herramientas "Ventana de Texto", "Ventana de Gráfico" y "Ventana de Tabla".



El análisis de página es rápido, sumamente preciso y tolera distorsiones: traza formas complejas e “irregulares”.



El análisis de página detectará incluso aquellas zonas en que hay **texto blanco sobre fondo negro**. El reconocimiento de dichos trozos no plantea problema alguno: la vista preliminar muestra el documento escaneado de forma correcta en la pantalla, pero Readiris “invierte” la imagen cuando es necesario para reconocer dichos bloques de texto. (Gracias al escáner, podrá generar imágenes *totalmente* invertidas de páginas que contengan texto blanco sobre fondo negro. Ver más adelante.)

UNA Y MEDIA, CLASIFICACIÓN DE VENTANAS

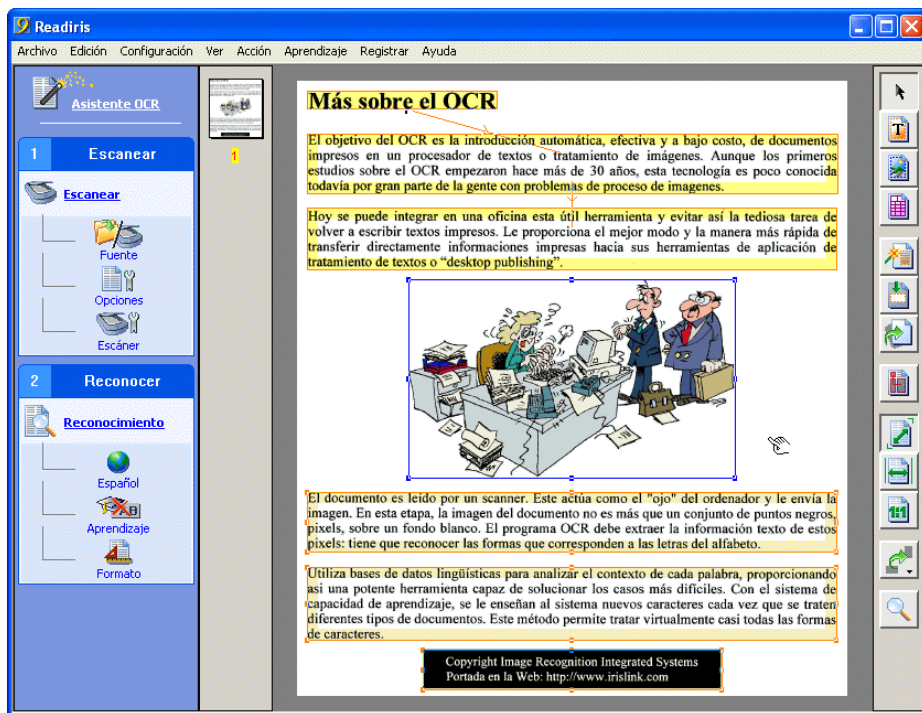
Readiris no sólo detecta los diferentes tipos de bloques, sino que también los *clasifica*: esta clasificación se realiza por defecto de arriba a abajo y de derecha a izquierda, con objeto de poder trabajar con documentos que contengan columnas.



Evidentemente, puede usted modificar el **orden de clasificación**. Para ello, haga clic en el botón "Ordenar" de la barra de herramientas de imágenes. El puntero del ratón se convierte en una manecilla en cuanto se activa el "modo de clasificación".



Haga clic en las ventanas que desee incluir. Las ventanas en las que *no* haga clic no serán clasificadas, quedando excluidas del reconocimiento. Es muy fácil saber qué ventanas han sido seleccionadas y cuáles no: las ventanas seleccionadas se visualizan con todo su color, mientras que las no lo han sido aparecen difuminadas.



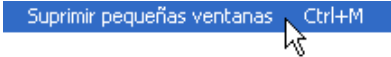
Por defecto, el análisis de página está activado. Para obligar a Readiris a descomponer la página actual - debido a que desactivó el análisis de página accidentalmente o borró algunas ventanas por error y desea volver a efectuar el análisis de página - simplemente haga clic en el botón "Analizar Página" de la barra de herramientas de imágenes.





Seleccione el idioma del documento *antes* de ejecutar el análisis de página siempre que vaya a reconocer documentos asiáticos. Para estos idiomas se utilizan rutinas específicas, dado que por una parte, el espaciado interlineal de los documentos de lenguas asiáticas es, por regla general, mayor que el de los documentos de lenguas “occidentales”. Además, el texto está formado por pequeños gráficos denominados “ideogramas” que podrían ser fácilmente confundidos con zonas gráficas en los documentos “occidentales”. Y por último, el texto puede leerse de arriba a abajo y de derecha a izquierda. Y si ha olvidado de seleccionar el idioma adecuado, selecciónelo después. Readiris volverá a ejecutar el análisis de página automáticamente.

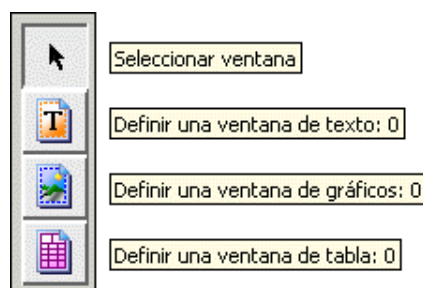
Algunos documentos presentan pequeñas manchas sobre la página, pueden generar un borde negro alrededor de la imagen real, etc. Para eliminar todas las pequeñas ventanas - suponiendo que no contienen texto - haga clic en el mando "Borrar Pequeñas Ventanas" bajo el menú "Edición".



Suprimir pequeñas ventanas Ctrl+M

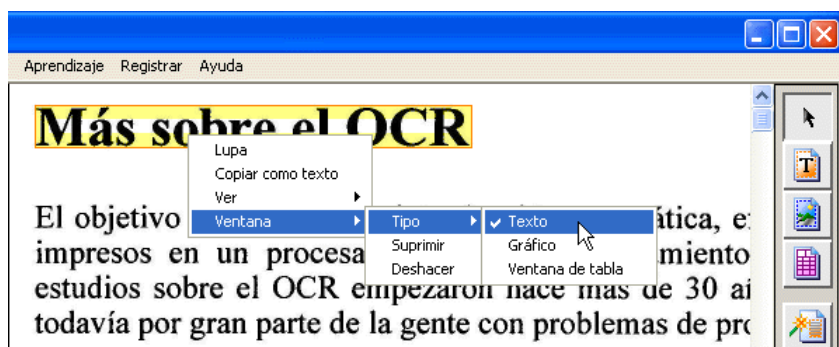
DOS, TRAZAR VENTANAS MANUALMENTE EN UNA IMAGEN

El análisis de página es la forma automática de trazar ventanas en una página escaneada. También puede dividir manualmente una imagen con las **herramientas de trazado de ventanas** de Readiris.



Para **trazar** un rectángulo alrededor de zona concreta, seleccione la herramienta correspondiente en la barra de herramientas de imágenes, coloque el cursor en la parte superior izquierda de la ventana que desee trazar, pulse el botón del ratón, extienda la ventana moviendo el ratón hacia la esquina inferior derecha de la ventana y vuelva a pulsar el botón del ratón. (No se permiten lados inferiores a 1 mm., pues no podrían contener ni un solo carácter.)

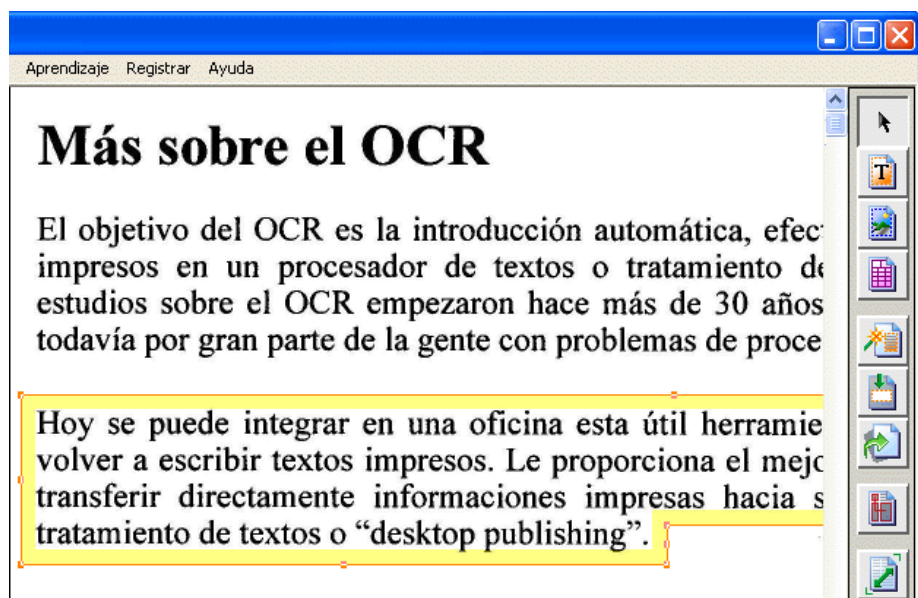
En caso de haber seleccionado el tipo de zona erróneo, no se preocupe: podrá cambiarlo rápidamente haciendo clic con el botón derecho del ratón sobre una ventana y seleccionando el mando "Ventana - Tipo" desde el menú contextual.



Las ventanas se clasifican automáticamente en el orden de su creación: una flecha indican el orden de clasificación.



También puede enmarcar bloques “irregulares” de texto trazando **ventanas poligonales** alrededor de ellos. Las ventanas no rectangulares se crean fusionando zonas rectangulares: tan pronto como dos rectángulos (del mismo tipo) se intersectan, se convierten automáticamente en una sola ventana. De cierta forma, es como construir una casa añadiendo una habitación tras otra... (Crear ventanas de tablas poligonales carece de sentido.)



Además, puede combinarse el trazado manual de ventanas con la clasificación de éstas: puede trazar nuevas ventanas incluso cuando esté activado el “modo de clasificación”. Así, puede realizar una clasificación automática para incluir las ventanas detectadas y después crear otras ventanas manualmente allí donde el análisis de página no ha ofrecido los resultados esperados. ¡En cuanto comience

el trazado manual de ventanas, con el “modo de clasificación” activado, se borrarán todas las ventanas que no haya seleccionado!

Para modificar, mover y eliminar ventanas, primero debe **seleccionarlas**. Para ello, seleccione la herramienta "Selección de Ventana" (la “flecha”) en la barra de herramientas de imágenes y haga clic en el interior de la ventana. Aparecerán unas marcas rectangulares en cada esquina y en el centro de los lados de la ventana.



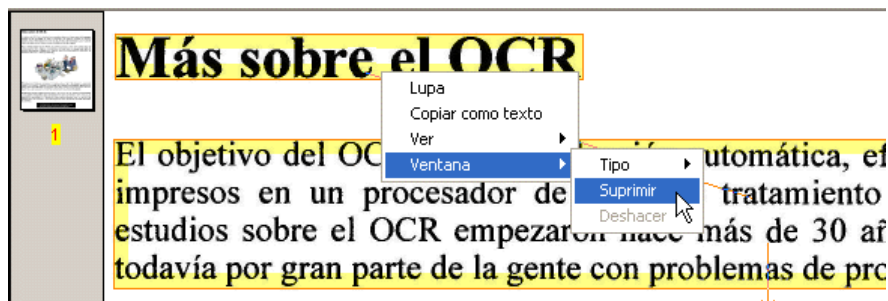
Más sobre el OCR

Para **deseleccionar** ventanas, pulse el botón del ratón en cualquier otro lugar. Para seleccionar **ventanas adicionales**, mantenga pulsada la tecla Mayús mientras hace clic en dichas ventanas. Para seleccionar una ventana y las **ventanas incluidas** (de otro tipo), mantenga pulsada la tecla Ctrl y haga clic en la ventana principal.

Hasta aquí la selección de ventanas. Para **modificar** una ventana de texto, selecciónela, coloque el cursor del ratón sobre una marca y arrastre el lado para cambiar el tamaño de la ventana.

Para **mover** una ventana, simplemente selecciónela y arrástrela al lugar deseado.

Para **eliminar** ventanas, seleccione la(s) ventana(s) y elija el mando "Borrar" del menú "Contexto". Dicha función borra la(s) ventana(s) seleccionada(s) y la que se encuentre bajo el cursor del ratón. Al pegar ventanas, éstas se insertan en su posición original y debe arrastrarlas al lugar deseado.



De igual modo, puede seleccionar zonas y optar los mandos "Cortar" o "Borrar" del menú "Edición". El mando "Cortar" corta la(s) ventana(s) y la(s) envía a una memoria temporal interna; la función "Borrar" borra irreversiblemente la(s) ventana(s). Al pegar ventanas, éstas se insertan en su posición original y debe arrastrarlas al lugar deseado.

De hecho, *todos* los mandos del menú "Edición" se aplican a las ventanas: puede eliminarlas, cortarlas, copiarlas y pegarlas. También puede utilizar la opción "Deshacer": si por error ha eliminado, movido, cambiado de tamaño, etc. algunas ventanas, "Deshacer" anulará la última operación.

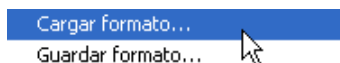
Edición		
Deshacer		Ctrl+Z
Cortar		Ctrl+X
Copiar		Ctrl+C
Pegar		Ctrl+V
Borrar		Delete
Suprimir pequeñas ventanas		Ctrl+M
Seleccionar todo		Ctrl+A

Existen macros para todas las opciones. He aquí un ejemplo: para borrar todas las ventanas presentes, puede elegir la opción "Seleccionar Todo" del menú "Edición" o la macro equivalente Ctrl+A y hacer clic en el mando "Borrar" del

menú "Edición" o su macro Supr. Ahora ya está listo para reproducir el diseño necesario. Para restaurar el diseño anterior, elija "Deshacer" o su macro Ctrl+Z.

TRES, GUARDAR MODELOS DE VENTANAS

Los diseños de ventanas resultantes pueden guardarse en forma de **modelos de zonas** para futuro uso, con la opción "Guardar Formato" del menú "Archivo" y cargarlos en memoria con la opción "Cargar Formato".



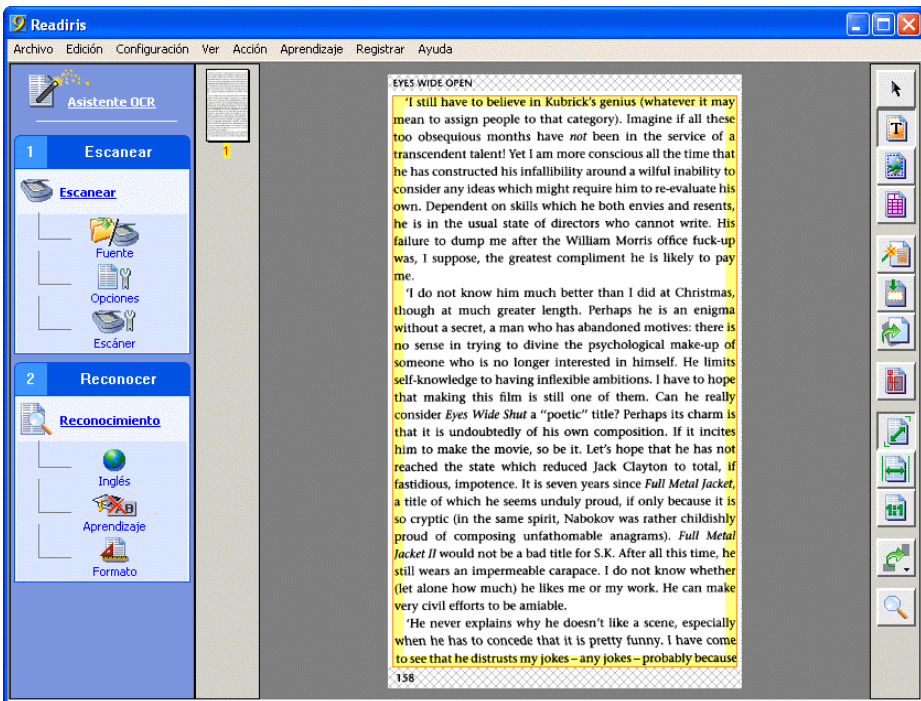
Si desea reconocer documentos con un diseño similar, por ejemplo, un informe de 50 páginas en la que el encabezado y el pie de página deben excluirse por razones obvias, se puede aplicar un modelo único para dividir en zonas las 50 páginas.

Al cargar un modelo en memoria se desactiva automáticamente el análisis de página. La división en zonas del modelo se mantiene activa hasta que vuelva a activar el análisis de página en la barra principal de herramientas.

De hecho, existe una eficaz alternativa para establecer las plantillas de determinación de zona. La herramienta de vista preliminar "Ignorar la Zona Exterior" limita la descomposición de la página a la sección “cortada” de la imagen.



Seleccione esta herramienta y encuadre la sección de la imagen que quiere procesar. Cuando está trabajando con un documento con múltiples páginas, puede eliminar la misma zona exterior mediante un análisis de página de cada página. (Vuelva a realizar el análisis de página para cancelar el “corte” de la imagen o cambie las zonas manualmente.)



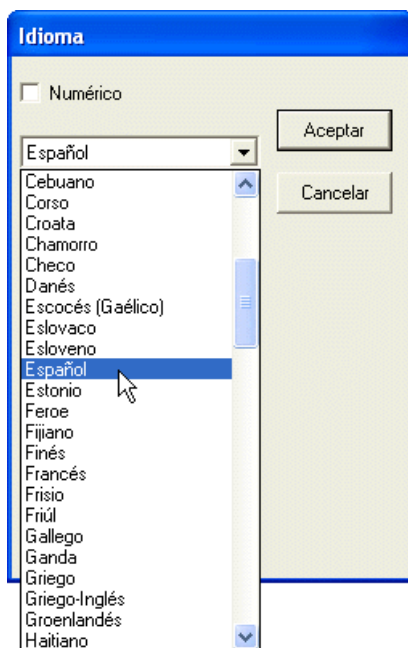
READIRIS LO LLEVA POR TODO EL MUNDO

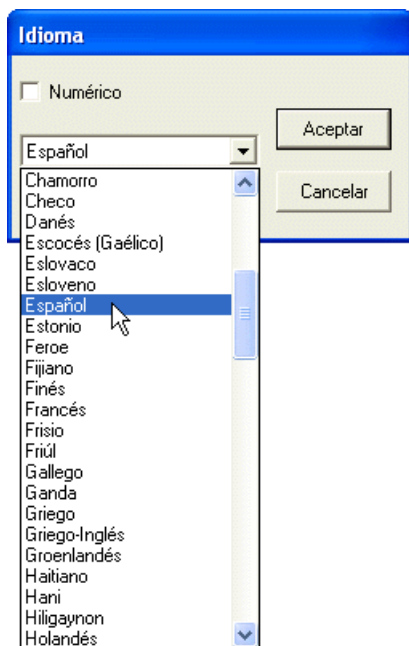
Suponiendo que las ventanas estén correctamente definidas, ahora ya casi está listo para ejecutar el reconocimiento de caracteres. Decimos “casi” porque aún no hemos verificado el idioma y las propiedades del documento.

Las propiedades de idioma se encuentran en la barra principal de herramientas.



Haga clic en el botón "Idioma" para modificar el idioma del documento.





Para pasar directamente a un idioma pulse la letra inicial de éste: si está seleccionado el español y desea localizar el danés, pulse la letra "D" para ir directamente hasta él. Si varios idiomas tienen la misma inicial, pulse la letra varias veces para recorrer las opciones. Por ejemplo, Readiris lee waray y wolof. Si pulsa "W" una vez, seleccionará el waray, si vuelve a pulsar "W" seleccionará el wolof, y pulsando "W" una tercera vez volverá al waray. (Para ir a *otra* letra, por ejemplo la "T", pulse primero la tecla Retroceso.)

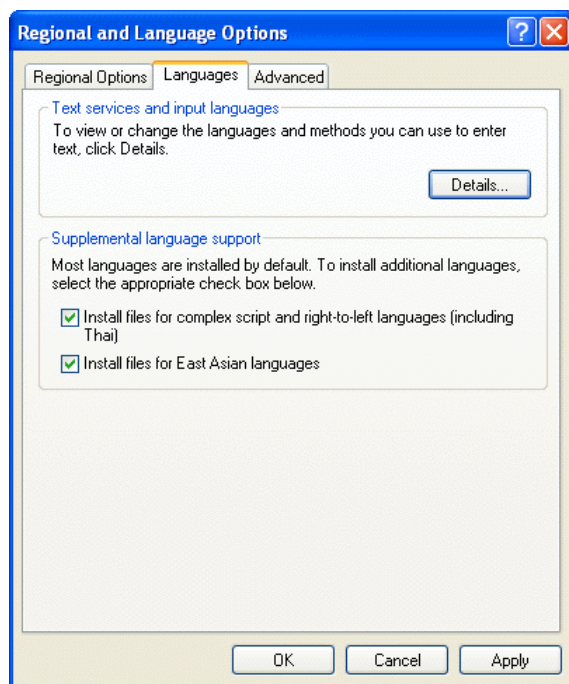
Readiris no se limita ni con mucho al español: se pueden utilizar hasta 107 **idiomas**: se pueden utilizar todos los idiomas americanos y europeos, incluyendo los idiomas centroeuropeos, cirílicos (“rusos”), bálticos, el griego y el turco.

Opcionalmente, puede reconocer documentos escritos en **lenguas asiáticas**. El *software* opcional “Módulo OCR asiático” permite el reconocimiento del japonés, el chino simplificado, el chino tradicional y el coreo. (El chino simplificado se habla en el continente chino y en Singapur, y el chino tradicional lo utilizan Hong Kong, Taiwan, Macau y las comunidades chinas ultramarinas.)

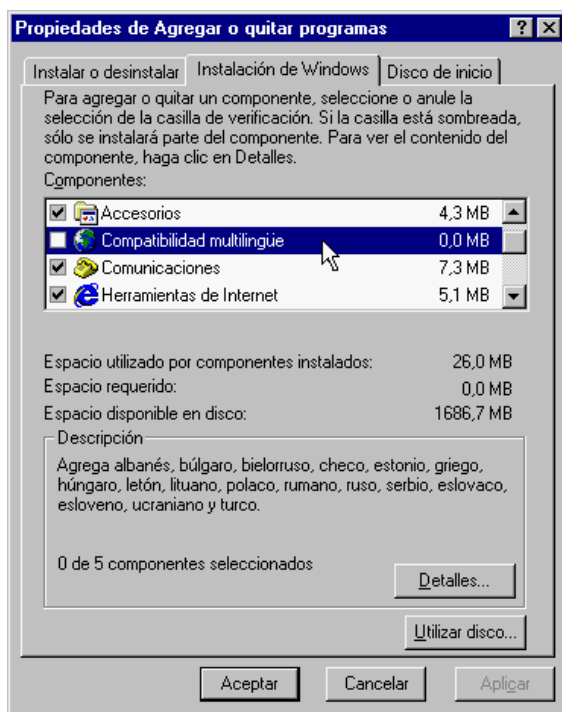
Cabe señalar que también se distingue entre las variedades británica y norteamericana - ¿o deberíamos decir internacional? - de la lengua inglesa.

Se requiere la configuración adecuada de Windows para mostrar en pantalla caracteres centroeuropeos, griegos, turcos, cirílicos y bálticos. Quizá deba instalar el componente de Windows "Compatibilidad Multilingüe" para que su sistema pueda procesar dichos idiomas.

En un sistema operativo Windows XP, 2000 y Windows NT 4.0, seleccione el icono "Opciones Regionales (y Idiomas)" en el menú "Panel de Control".



En un sistema operativo Windows ME y 98, seleccione el icono "Agregar o Quitar Programas" de "Panel de Control" para ver si el módulo "Compatibilidad Multilingüe" está instalado.



Para visualizar y editar documentos en lenguas asiáticas, deberá instalar una versión asiática del sistema operativo de Windows o utilizar un programa de “emulación” especializado (como, por ejemplo, UnionWay AsianSuite o TwinBridge AsianBridge) en una versión occidental de Windows, para representar correctamente los ideogramas de estas lenguas asiáticas. Podrá, asimismo, utilizar Word 2003, Word 2002 o Word 2000 para visualizar y editar este tipo de documentos, dado que Office 2003 System, Office XP y Office 2000 han sido específicamente diseñado para trabajar con documentos en muy diversos idiomas.



Consulte el **archivo “Léame”** de Readiris para obtener más información sobre este tema.

La selección del idioma adecuado para el documento es imperativa. En función de la selección del idioma, el programa sabe qué **juegos de símbolos** debe reconocer. El soporte multilingüe garantiza un reconocimiento correcto de caracteres “exóticos” como ç, ß, γ y ø.

Asimismo, el programa hace un extenso uso de **bases de datos lingüísticas** para convalidar sus resultados. Supongamos que se debe leer la palabra "presidente" y una mancha de tinta hace que la "r" parezca una "f". Al buscar en el léxico español, Readiris detectará de forma autónoma que está leyendo la palabra "presidente" y que no tiene sentido reconocer el símbolo "f". Por supuesto, esta **técnica de “autoaprendizaje”** depende en gran medida del contexto lingüístico.

A menudo, la lingüística resulta muy útil para resolver **casos ambiguos**, como una "O" que puede confundirse erróneamente por un '0'. Otro ejemplo es la letra "l" y el número '1' que tienen una forma idéntica en muchas fuentes; piense por ejemplo en los textos escritos con antiguas máquinas de escribir. El contexto lingüístico ayuda a determinar si se trata de una "l" o de un '1'.

La ilustración a continuación muestra varias formas de 'l' y '1'. Las formas de la primera línea carecen de ambigüedad, las formas de la segunda son ambiguas, pero la lingüística permite resolverlas. Si no basta con el contexto, el usuario debe intervenir.

193 1950s. 1hr
Well, Rossellini

READIRIS CAMBIA DE IDIOMA CUANDO ES NECESARIO

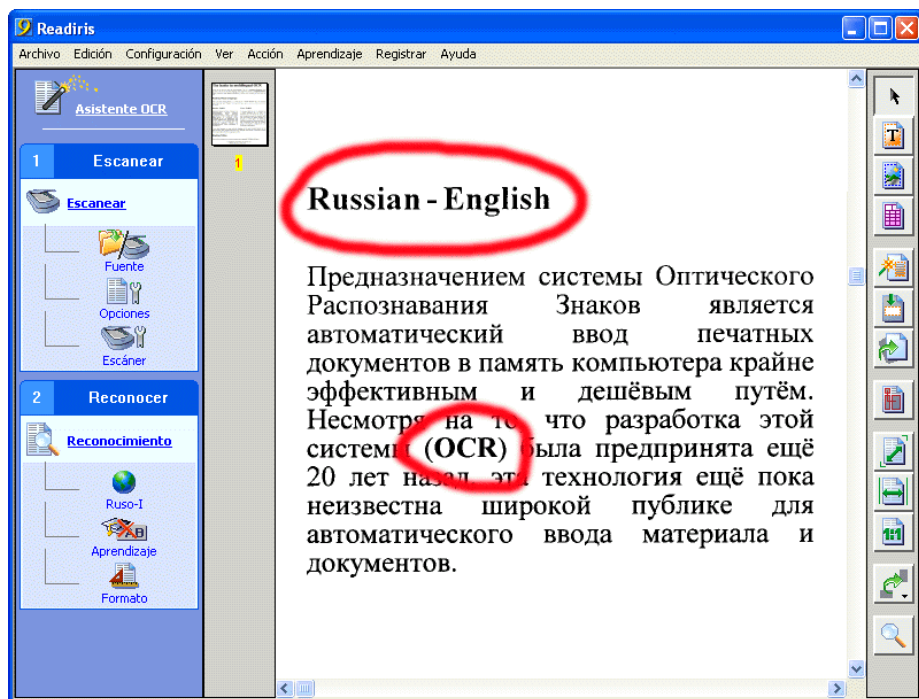
Pero eso no es todo: Readiris puede cambiar de idioma a media frase sin intervención del usuario. Cuando aparecen palabras occidentales en documentos escritos en caracteres griegos, cirílicos o asiáticos - muchos nombres propios no

modificables, marcas, etc. se escriben respetando su grafía latina -, Readiris puede adaptarse automáticamente al alfabeto apropiado. O sea, es posible activar simultáneamente un **alfabeto mixto** de griego, cirílico o asiático y latino.

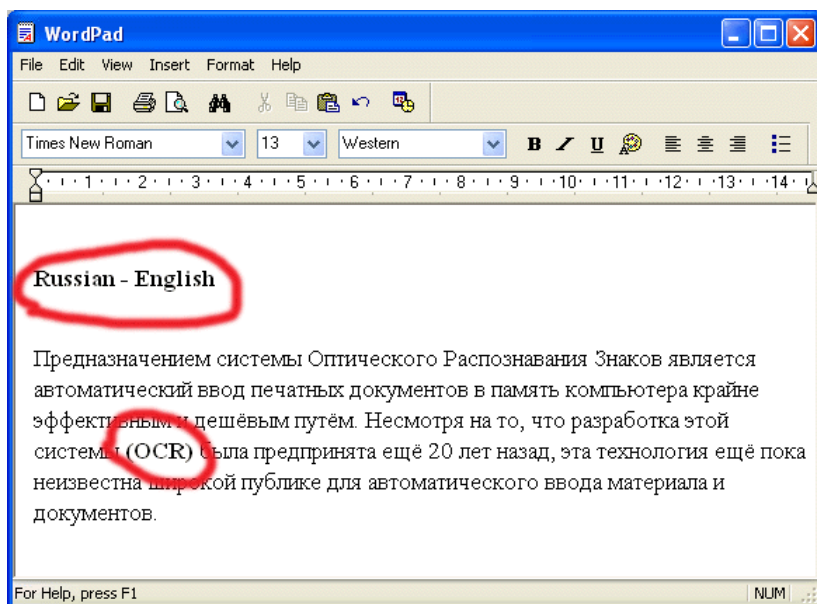
Asegúrese de seleccionar "Griego-Inglés" o la propiedad adecuada de cirílico - por ejemplo "Bielorruso-Inglés". En otras palabras, si selecciona únicamente "Griego" o "Bielorruso" como idioma del documento, ¡no espere que salgan bien los caracteres latinos!



He aquí un ejemplo en el que un texto contiene algunas palabras inglesas - abra el archivo de imagen ALPHABETS.TIF si desea intentarlo por sí mismo.



El resultado final es parecido a éste cuando se abre con el tratamiento de textos - quizá debe seleccionar una **fuentes** cirílica para mostrar correctamente el texto ruso en pantalla.



Para **mezclar otros idiomas**, simplemente seleccione el idioma con el mayor número de caracteres. Si tiene un documento en el que, digamos, una traducción inglés está junto al texto francés, debe seleccionar el idioma francés para garantizar que se reconozcan los caracteres acentuados como ç, é y ù.

DEFINICIÓN DE LAS CARACTERÍSTICAS DEL DOCUMENTO

Una vez seleccionado el idioma, pasaremos a las otras características del documento. Puede afinar el reconocimiento especificando algunas de las características del documento: tipo de fuente y espacio entre caracteres. (Estas



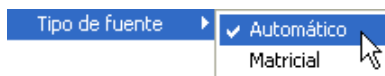
opciones no se aplican a los documentos asiáticos.) Expliquemos lo que esto significa.

Empecemos con el mando "Tipo de Fuente" en el menú "Configuración". El tipo de fuente distingue los documentos "normales" de los **documentos matriciales**. Los símbolos matriciales de calidad "borrador", también denominados "9 agujas", están formados por puntos aislados y separados y su reconocimiento requiere rutinas altamente especializadas.

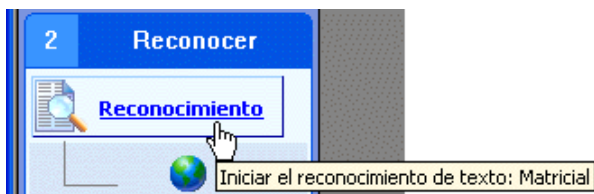
ape-descended life

La impresión con matriz de agujas "calidad de carta", también denominado "25 agujas", requiere la configuración "normal" al igual que las **calidades de impresión** imprenta, mecanografiado, impresión láser e impresión con chorro de tinta.

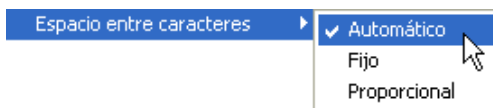
La opción "Automático" significa que Readiris detectará por sí mismo el tipo de fuente. Deje Readiris detectar automáticamente el tipo de los caracteres en todos los casos - ¡a no ser que este seguro que va a leer documentos matriciales! (Obviamente, "Automático" es el valor por defecto.)



La etiqueta emergente del botón "Reconocimiento" indica cuál es el tipo de letra activo. Cuando no se muestra ningún mensaje, se usará la calidad de impresión "normal"; cuando aparece el mensaje "Matricial", significa que está activado el modo de lectura para textos impresos con impresora matricial.



El **ancho de fuente** puede establecerse con la opción "Espacio entre Caracteres" en el menú "Configuración".



Con fuentes *fijas* (o fuentes “monoespacios”), todos los símbolos de la fuente tienen el mismo ancho. Una “i” ocupa el mismo espacio horizontal que una “w”, como sucede en esta frase. Piense en los documentos escritos con una máquina de escribir, en la que el carro se mueve una distancia fija para cada símbolo escrito.

Un ancho *proporcional* significa que un carácter depende de su forma. Los símbolos como “m” y “w” son más anchos y ocupan más espacio horizontal en una línea que caracteres “delgados” como “l” ó “j”. Virtualmente todos los libros, revistas y periódicos están impresos en ancho proporcional.

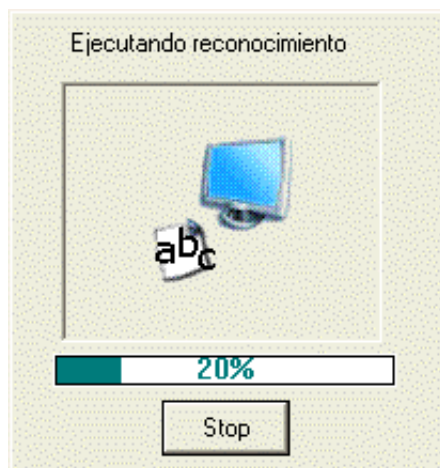
La solución más sencilla es dejar esta opción siempre en el valor por defecto “Automático”, que significa que Readiris detectará automáticamente el ancho de los caracteres.

¡READIRIS SE HACE CADA VEZ MÁS INTELIGENTE!

Una vez seleccionado el idioma del documento y establecidas las características del documento, active el aprendizaje interactivo y pulse el botón “Reconocimiento”.



El avance del reconocimiento se indica en la pantalla. Puede pulsar el botón “Stop” para interrumpir el reconocimiento de texto.



Al término del reconocimiento, Readiris entra en la fase interactiva de aprendizaje en la cual se activa el aprendizaje pulsando el botón "Aprendizaje" de la barra principal de herramientas.

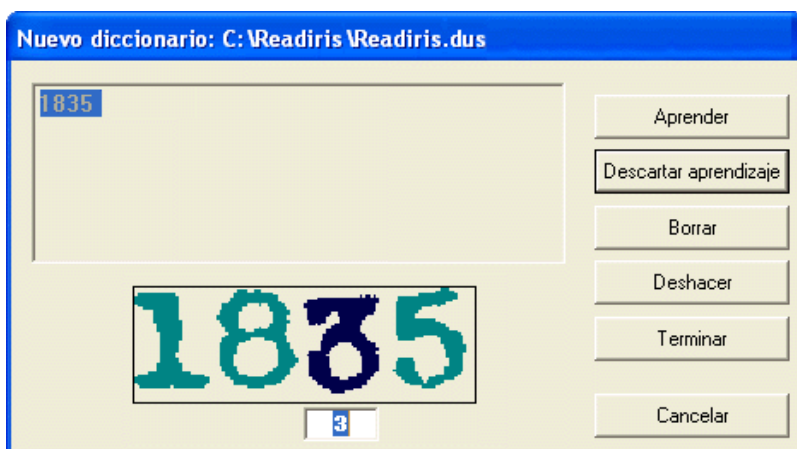
(El aprendizaje interactivo no está disponible para los documentos de lenguas asiáticas, pues carece de sentido en estos idiomas, que se componen de miles de símbolos diferentes - además habrían de poder introducirse los ideogramas mediante un teclado occidental, ¡cosa nada fácil!)

El **aprendizaje de fuentes** puede mejorar considerablemente la precisión del sistema de reconocimiento. Si el usuario intenta reconocer formularios distorsionados o desfigurados como los que se encuentran en documentos reales o bien formas estilizadas de fuentes que Readiris no reconoce de forma óptima, este aprendizaje puede resolver este "fallo" temporal.

El aprendizaje del usuario también se utiliza para que el sistema aprenda **símbolos especiales** que Readiris no es capaz de reconocer, como símbolos matemáticos y científicos y *dingbats*. Algunos ejemplos: Readiris puede aprender

a reconocer el símbolo " π " como "pi" o el *dingbat* "☎" como "Tel". (Sin embargo, la lista de símbolos reconocidos no puede extenderse a los símbolos " π " y "☎".)

El texto reconocido aparece progresivamente en pantalla y el sistema se detiene en los caracteres dudosos, o - si se trata de caracteres unidos ("ligaduras") - en cadenas de caracteres dudosas. Los caracteres dudosos siempre se presentan en su contexto y se destacan. Los caracteres no reconocidos se representan con una tilde.



Lo primero que debe hacer es comprobar si ha activado el diccionario de fuentes y el modo de diccionario adecuados - éstos se indican siempre en el título de la ventana de aprendizaje. Si no es así, haga clic en el botón "Cancelar" - la imagen del documento vuelve a aparecer con la división por zonas con la que se fue creado - active el diccionario de fuentes o el modo de diccionario adecuados y vuelva a ejecutar el reconocimiento. (Pronto se explicará el funcionamiento de los diccionarios de fuentes.)

De ser necesario, introduzca un carácter (o cadena de caracteres) en lugar de la forma incorrecta o desconocida y pulse uno de los siguientes botones.



Aprender

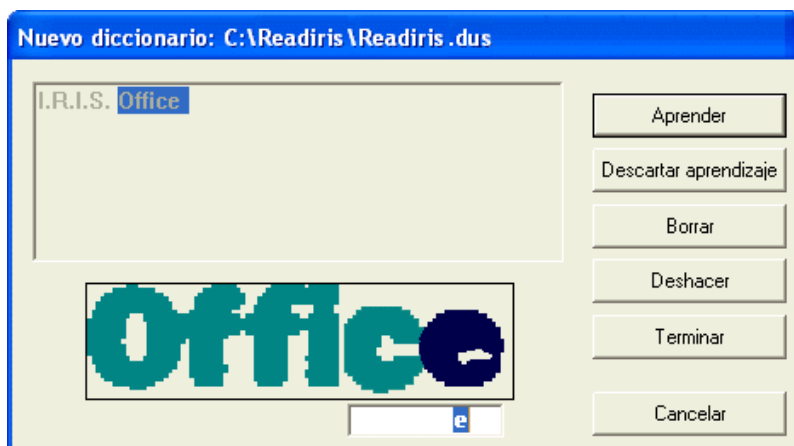
Está de acuerdo con la solución propuesta o la corrige. El programa guarda este carácter dudoso en el diccionario de fuentes como carácter “seguro”, definitivo. Los reconocimientos posteriores no requerirán su intervención, la forma se considerará aprendida de una vez por todas.

En el ejemplo anterior, el sistema se detiene en un carácter sucio y pulsamos "Aprender" para aceptar una forma que no puede confundirse con otros caracteres.

Descartar Aprendizaje

Está de acuerdo con la solución propuesta o la corrige. La diferencia respecto al botón "Aprender" consiste en que el símbolo aprendido obtiene el calificativo de “inseguro” en el diccionario. Para su reconocimiento futuro, el sistema propondrá la solución “aprendida”, pero aún requerirá una confirmación.

Este botón se utiliza para aquellos símbolos que pueden confundirse con otros: una "e" desfigurada que puede confundirse con una "c", una "t" dañada que puede parecerse a una "r", etc.



La "e" que se muestra anteriormente está muy dañada - de hecho es muy parecida al símbolo "o", por lo que debe pulsar el botón "Descartar Aprendizaje" a fin de no confundirlo con el símbolo "o".

Borrar

La forma que aparece en pantalla se elimina del resultado. Este botón se utiliza para ignorar el “ruido” presente en los documentos - puntos, manchas de café, etc. - que pueden reconocerse como puntos, comas y otras cosas, así como para borrar los símbolos no deseados.

Deshacer

Para volver y corregir errores. Puede deshacer las últimas 32 decisiones.

Terminar

El proceso de aprendizaje concluye, pero el reconocimiento de caracteres continua en modo automático. En lo sucesivo, todas las decisiones adoptadas por el usuario son aceptadas sin confirmación por parte del usuario.

Pulse este botón cuando vea que el reconocimiento es sumamente preciso y no requiere una corrección detallada.

Cancelar

No confunda el botón "Terminar" con el botón "Cancelar". Con "Cancelar" no se genera un resultado y vuelve a comenzar desde el principio, con "Terminar" se crea el texto, sólo que no ha sido corregido en sus detalles.



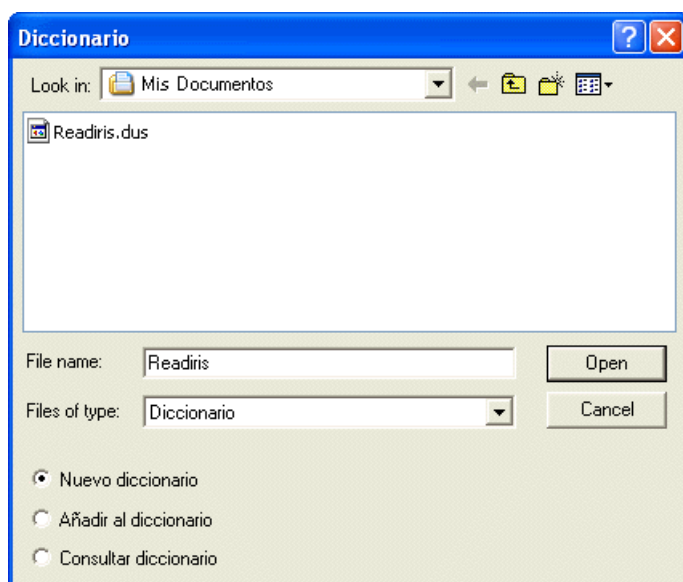
PAPEL DE LOS DICCIONARIOS DE FUENTES

Los resultados de cada sesión de aprendizaje se conservan temporalmente en la memoria del ordenador, pero pueden y deben guardarse en archivos denominados “diccionarios” para su uso ulterior.

(¡No confunda los diccionarios de fuentes con los léxicos (del usuario)! Los diccionarios de fuentes contienen formas de caracteres adquiridas durante la fase interactiva de OCR, mientras que los léxicos son bases de datos lingüísticas que ayudan al reconocimiento.)

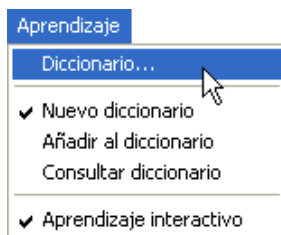
Debe cargar en memoria estos diccionarios de fuentes si desea reconocer documentos similares para aprovechar la información adicional que contienen; de esta forma, Readiris toma en consideración la información almacenada en estas bibliotecas de fuentes. ¡Podría decirse que Readiris se vuelve más inteligente cada vez que la usa!

¿Cómo funciona esto? El funcionamiento de los diccionarios de fuentes está controlado por el menú "Aprendizaje": tiene que seleccionar un diccionario con el mando "Diccionario" y determinar su modo de funcionamiento.



Los **diccionarios** de fuentes están limitados a 500 formas y le recomendamos que elabore distintos diccionarios para cada aplicación, por ejemplo, por tipo de documento. Por defecto, los diccionarios tienen la extensión *.DUS. El aprendizaje carece de efectos si el diccionario está lleno: los resultados no se conservan en la memoria ni se transcriben en el diccionario.

Dentro de la opción "Diccionario" o directamente en el menú "Aprendizaje" puede seleccionar un modo de diccionario. Existen tres modos de diccionario: nuevo, añadir y de lectura.



Si selecciona "Nuevo Diccionario", indicará que los resultados del aprendizaje deben guardarse en un *nuevo* diccionario. (Si selecciona un diccionario existente, su contenido se borrará.)

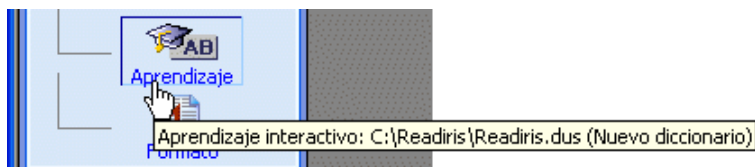
El modo “añadir” indica que los resultados del aprendizaje se guardarán en diccionario *existente*: el reconocimiento utiliza la información que ya contiene el diccionario y se le añaden nuevas formas de fuentes. En otras palabras, esta opción le permite elaborar un diccionario de fuentes en varios pasos.

(Si introduce un nombre de archivo para un nuevo diccionario y activa el modo “añadir”, se creará un diccionario de fuentes vacío y deberá completarlo.)

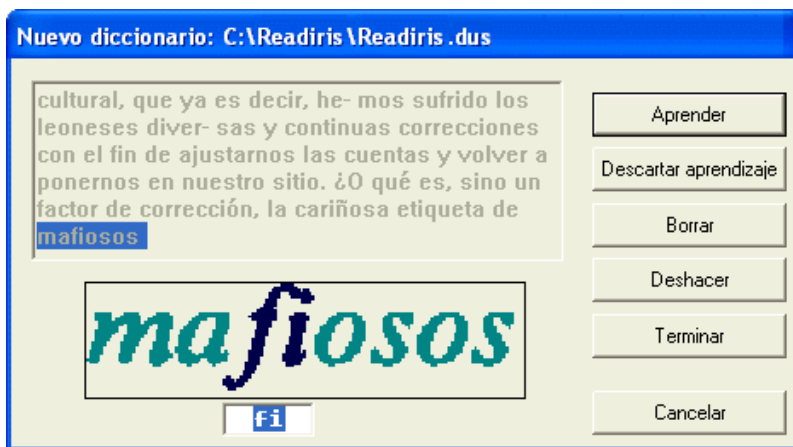
Con la última opción "Consultar Diccionario", el diccionario funciona en modo de “sólo lectura” y se aprovecha el diccionario *sin* añadirle nuevas formas.

Seleccione el modo nuevo si desea reconocer una sola página. Para reconocer muchas páginas del mismo tipo - con las mismas fuentes y la misma calidad de impresión - seleccione el modo “nuevo” para la primera página, el modo “añadir” para unas cuantas páginas más y el modo “de lectura” para el resto del documento o documentos.

La etiqueta emergente del botón "Aprendizaje" siempre indica el diccionario de fuentes que está activado y el modo en que opera éste.



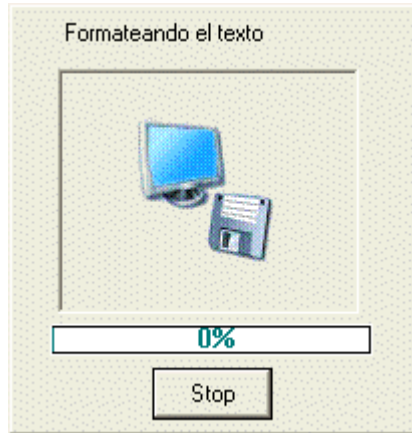
Al entrar en aprendizaje interactivo, el diccionario y su modo de funcionamiento se indican en el título de la ventana; pulse el botón "Cancelar" y vuelva a comenzar si éstos son incorrectos.



ENVÍO DE RESULTADOS A LA APLICACIÓN DE TRABAJO

Con el aprendizaje interactivo concluye el reconocimiento de caracteres. Si Microsoft Word es el programa receptor de datos por defecto, el procesador de textos se ejecuta automáticamente (si fuese necesario) una vez concluido el reconocimiento, y se inserta el texto reconocido.

Puede que en pantalla se muestre una barra de progresión del reconocimiento. (Esto dependerá de del tamaño del documento y de la complejidad del formato que haya de realizarse.)

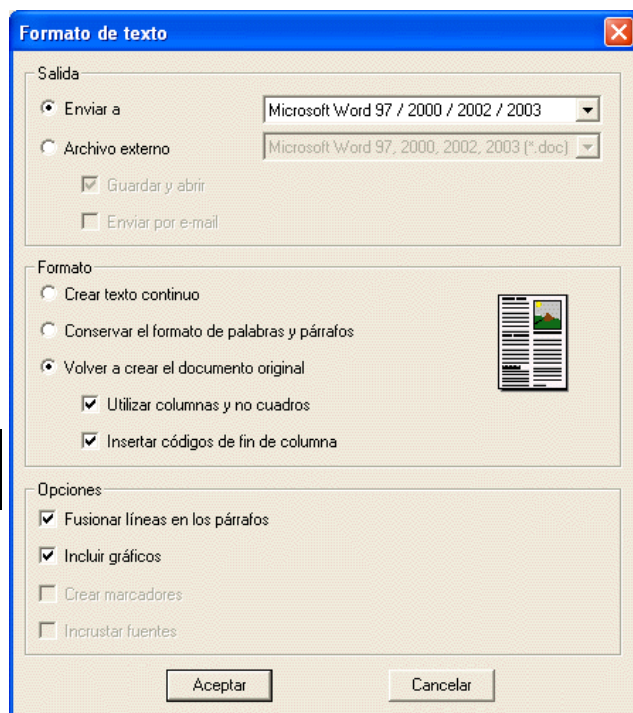


La imagen escaneada se visualiza de nuevo una vez efectuada la división en zonas para su posterior procesamiento y permanece en su sitio hasta que se escanee una nueva imagen.

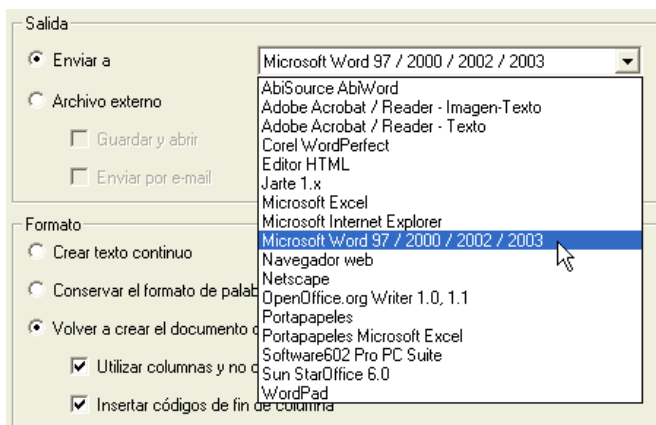
De hecho, ha convertido un documento en papel en un archivo digital modificable, ¡y en un tiempo 40 veces inferior al que habría empleado si hubiera tenido que volver a mecanografiarlo! Ahora compárelo con la imagen que tiene en la ventana de Readiris.

En realidad, Readiris le ofrece tres métodos diferentes a la hora de guardar el resultado del reconocimiento: por un lado, enviar directamente el documento reconocido a una aplicación. Por otro, guardar el resultado en un archivo externo y, por último, copiar el resultado en el portapapeles de Windows.

El **destino de salida** se selecciona con el botón "Formato" de la barra de herramientas principal (o con la opción "Formato Texto" del menú "Configuración").



La función "Enviar a" permite un vínculo directo de OCR entre el escáner y las aplicaciones para Windows. ¡Con ella podrá **enviar** directamente el documento que acaba de escanear a un procesador de textos, hoja de cálculo o programa de navegación, a Adobe Reader, etc.!



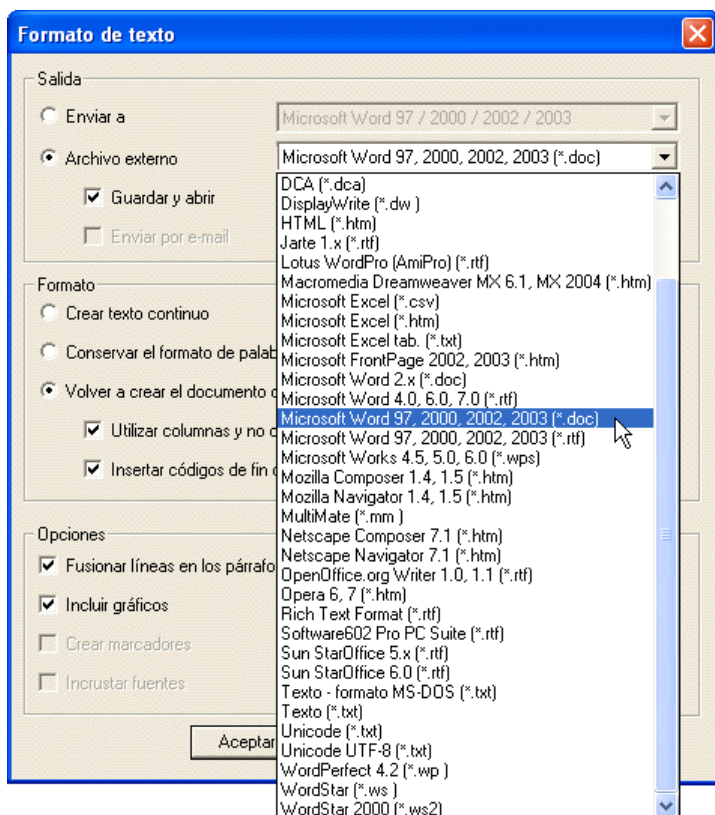
Una vez concluido el reconocimiento, se ejecuta la aplicación de destino y el documento reconocido se visualiza en un nuevo archivo de texto o en un libro de trabajo.



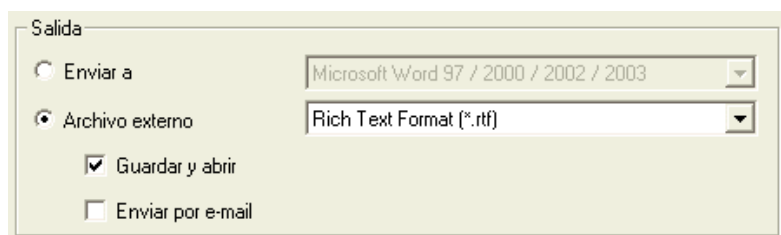
No olvide que la función "Enviar a" también permite enviar el texto reconocido al **portapapeles** de Windows, de forma que no es absolutamente imprescindible exportar el resultado de la lectura, ni guardarlo en un archivo externo...

GUARDAR LOS RESULTADOS EN UN ARCHIVO DE TEXTO

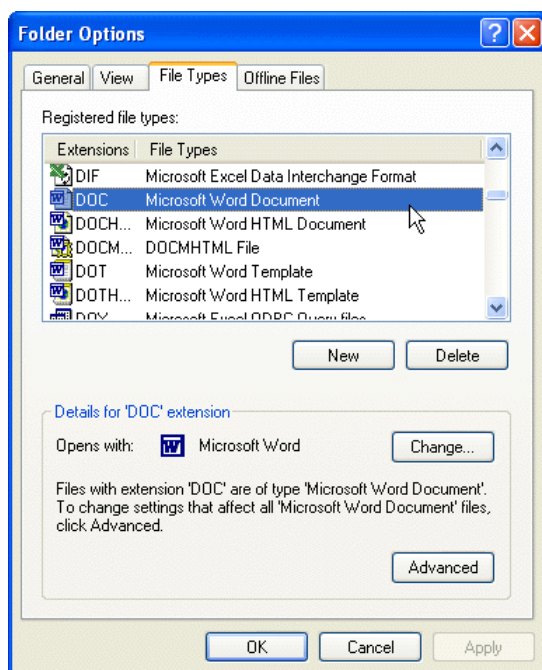
En efecto puede guardar el resultado del OCR en un archivo “externo”. Aquí también Readiris reconoce una gran variedad de los formatos de archivo ofrecidos por los procesadores de texto, hojas de cálculo y aplicaciones web más conocidos.



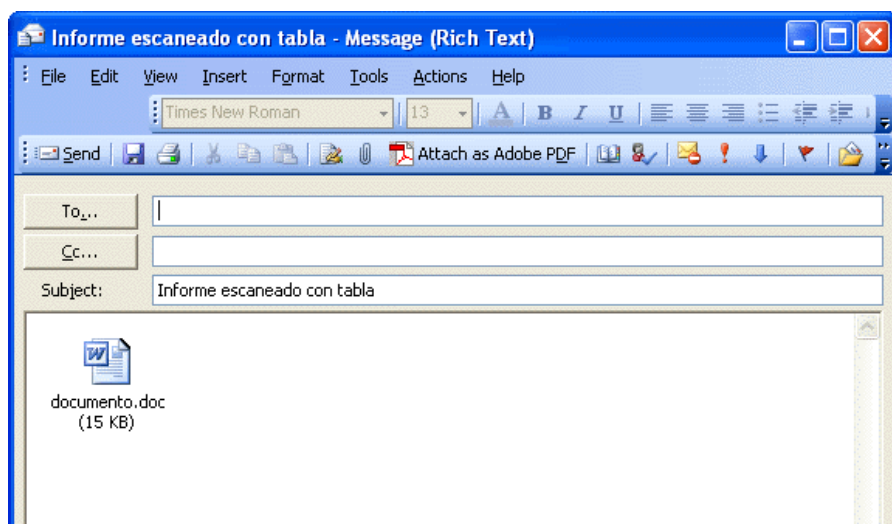
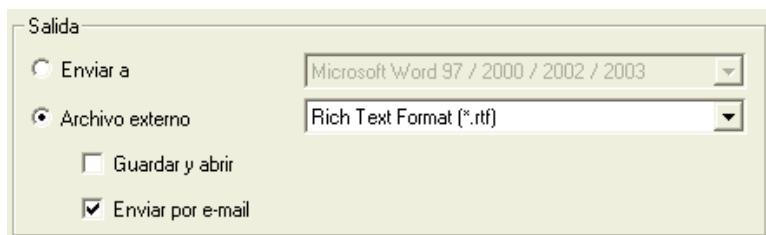
La opción "Guardar y Abrir" es muy parecida a la función "Enviar a", puesto que le permite abrir el documento reconocido cuando ya se ha guardado.



Sin embargo, el método usado para dirigir la aplicación deseada es distinto. En esta ocasión, los **tipos de archivo de Windows** determinan qué aplicación se iniciará. Es como si hiciera doble clic sobre el archivo de salida desde el Explorador de Windows... (Con la opción "Enviar a", Readiris dirige directamente las aplicaciones específicas deseadas.)

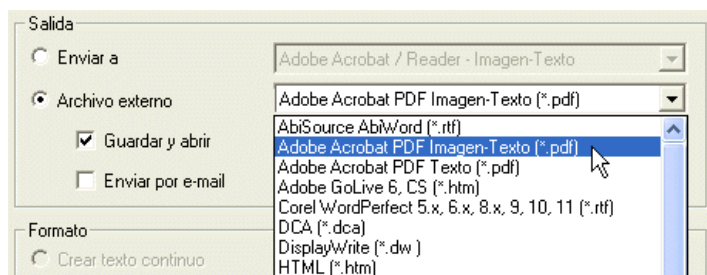
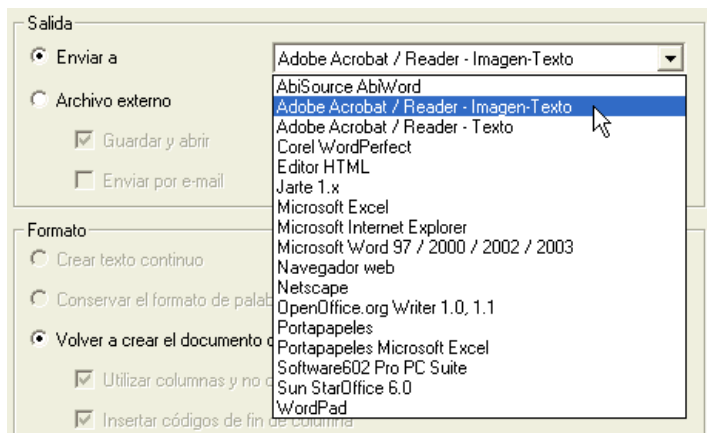


La opción "Enviar por E-mail" crea un nuevo mensaje de **correo electrónico** y añade el documento reconocido como archivo adjunto del mensaje. ¿Conoce un método más rápido para distribuir un documento en papel...?

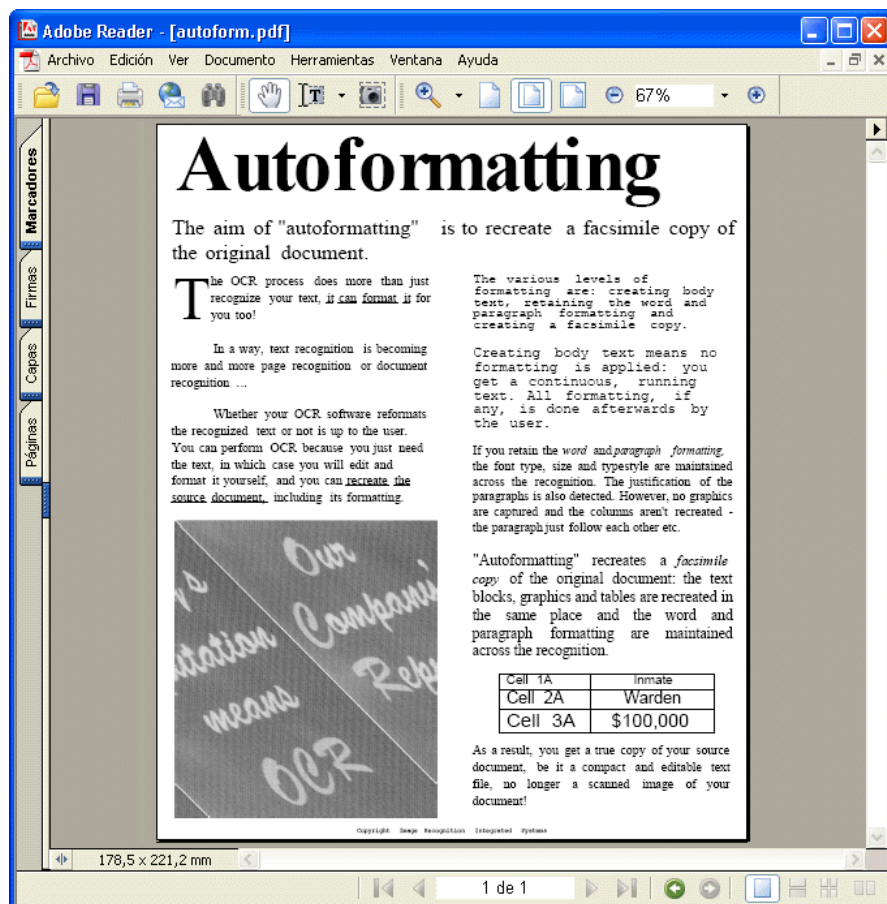


CREACIÓN DE DOCUMENTOS TRANSPORTABLES

Veamos más atentamente uno de los formatos reconocidos: **Acrobat Reader PDF**. Readiris le permite crear documentos PDF de dos tipos - PDF Texto y PDF Imagen-Texto.



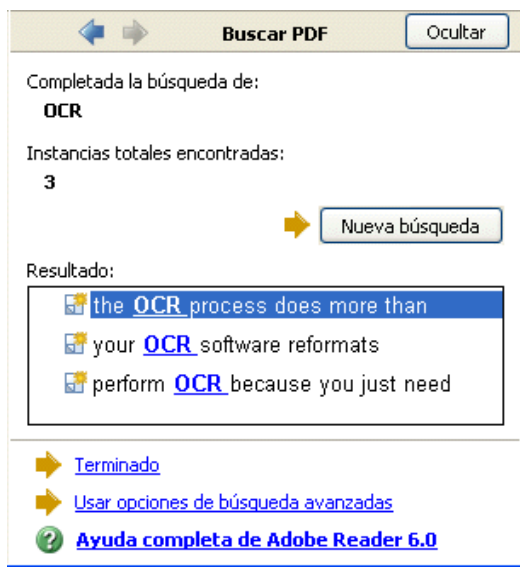
¿Cuál es la diferencia entre los dos? Cuando selecciona el formato "PDF Texto", Readiris crea un fichero PDF que contiene el resultado del texto. (Es posible que haya gráficos pero sólo cuando aparecen zonas gráficas en la página - fotografías, ilustraciones, etc.) Es decir que la imagen de la página *no* está contenida en el documento PDF, compuesto de un solo nivel.



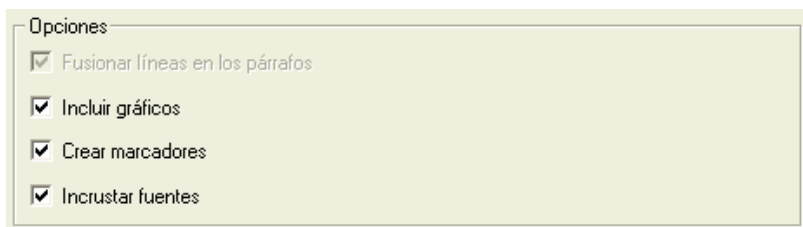
El formato "PDF Imagen-Texto" genera varios resultados: Readiris crea un fichero PDF en lo cual puede efectuar una búsqueda, que contiene el texto



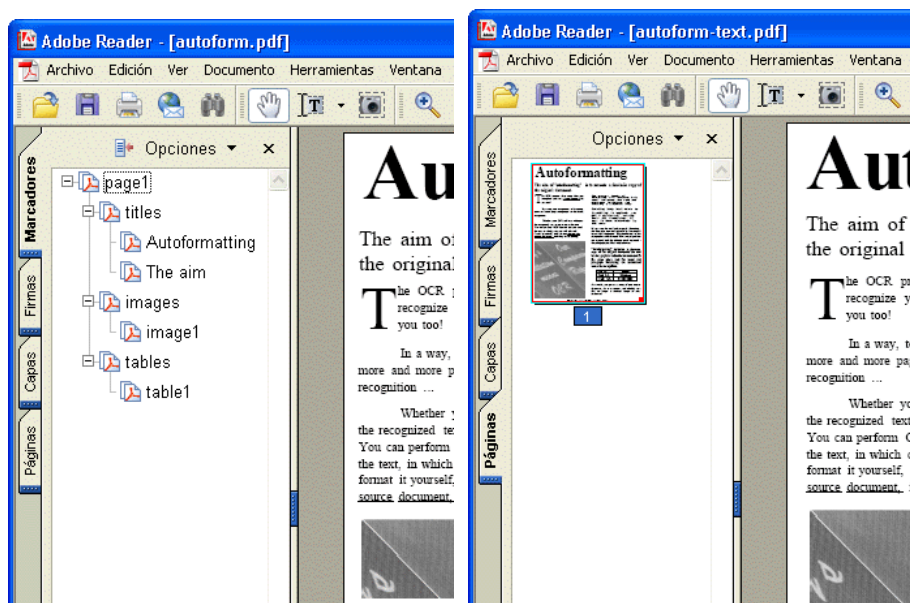
reconocido y la imagen de la página. La imagen de la página está contenida encima del texto en un documento PDF compuesto de dos niveles. ¡Utilice la opción "Buscar" de Adobe Reader y le aparecerá evidente!



Haga clic en el botón "Formato" para ver dos opciones relativas al formato PDF Acrobat: "Crear Marcadores" e "Incrustar Fuentes".



La opción "Crear Marcadores" permite crear un **marcador** para cada elemento del documento - los gráficos tanto como las zonas de texto y tablas. Para las zonas de texto, Readiris aplica un algoritmo inteligente que crea un título, un "externo" para cada zona; las tablas y gráficos llevan simplemente un número. (Otro elemento de navegación de los documentos PDF, las **miniaturas** de página, se pueden crear de manera dinámica con el programa Adobe Reader.)



La opción "Incrustar Fuentes" inserta fuentes en los archivos PDF. Al incrustar fuentes, se evita la sustitución de las mismas cuando el lector examina e imprime el documento reconocido. Asegura que el lector - cualquiera que sea la configuración de su ordenador - vea el texto con las fuentes originales. Sin embargo, al insertar fuentes se aumenta (un poco) el tamaño de los documentos reconocidos.



... O RECONOCIMIENTO DE DICHOS DOCUMENTOS

Veámoslo en el otro sentido. Dado que Readiris ofrece un reconocimiento completo del formato PDF Adobe Acrobat, no sólo le permite generar archivos PDF, sino que también puede *leerlos*!

Puede que la **conversión de los documentos PDF** sea una de las principales aplicaciones de Readiris. Existen diversas razones que lo justifican. En primer lugar, es una manera de convertir imágenes en texto: ya que los documentos PDF con imágenes abiertos ejecutan el reconocimiento y guardan el resultado del OCR en un documento de texto (en cualquier formato de texto reconocido). Los archivos de texto se pueden editar, mientras que los archivos con imágenes no son editables.

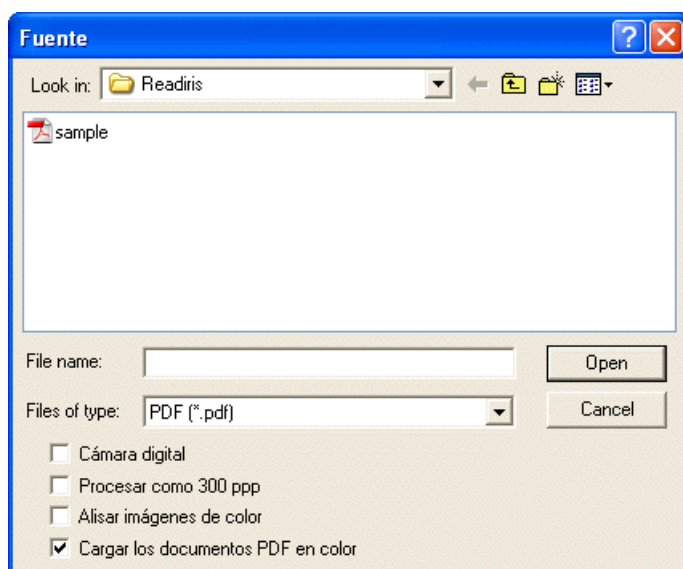
En segundo lugar, le permite convertir archivos PDF con imágenes en archivos PDF de texto. Para ello debe ejecutar el reconocimiento de archivos PDF con “sólo imágenes” y guardar los resultados del OCR... ¡como documentos PDF de texto! Los archivos PDF de texto se pueden buscar y editar, mientras que con los archivos PDF con “sólo imágenes” esto no es posible.

Finalmente, la conversión de archivos PDF es una manera de “desbloquear” el contenido del PDF. Piense en los documentos PDF de “sólo lectura”, cuyo texto normalmente es inaccesible. Con los archivos PDF desprotegidos, se puede recuperar el contenido (copiar y guardar en un archivo texto), mientras que el contenido de los archivos “sólo lectura” no se puede extraer. Dichos documentos sólo se pueden ver e imprimir.

Un “detalle” importante es que Readiris no abre documentos PDF protegidos con una contraseña, ¡ni siquiera en el caso de que Readiris haya roto el resto de barreras de seguridad PDF!

Proceda del modo habitual: cargue los archivos PDF en la memoria al tiempo que abre imágenes previamente escaneadas - faxes, instantáneas realizadas con cámara digital, etc. Haga clic en el botón "Stop" o Escape para interrumpir el proceso de descarga entre dos páginas.

Sin embargo, existe una opción específica relacionada con los archivos PDF, que permite abrirlos como documentos en color o en blanco y negro. ¡Se ofrece esta opción porque explorar documentos en color es mucho más lento!



RECONOCIMIENTO DE VARIAS PÁGINAS

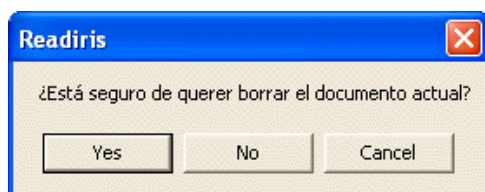
Después del OCR, la imagen escaneada se visualiza de nuevo con la división en zonas para su posterior procesamiento.

A partir de este momento, puede abrir el texto reconocido con su tratamiento o editor de textos, importarlo a un programa de maquetación (DTP), o a cualquier otras aplicaciones de texto. Compárelo ahora con la imagen que tiene en la ventana de Readiris.

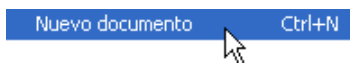


Pero ¿cómo se guarda el texto de las páginas adicionales? En otras palabras, ¿cómo se procesan los documentos compuestos de varias páginas? Es muy sencillo: vaya ejecutando las páginas reconocidas y guarde los resultados en el mismo archivo. (Asegúrese de que el documento no está abierto para poder escribir en él.) No olvide tampoco activar el modo “añadir” en el diccionario de fuentes, de forma que pueda continuar cómodamente con el aprendizaje de fuentes.

Conforme vaya escaneando páginas (o abriendo archivos gráficos), tendrá que decidir si quiere iniciar un nuevo documento o completar el que está abierto en ese momento.

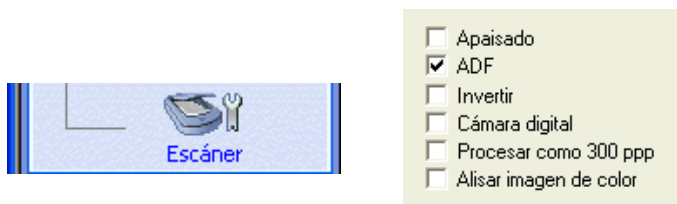


Si responde "No", seguirá añadiendo páginas al documento abierto; si responde "Sí", creará un nuevo documento. Esta última respuesta corresponde a la función "Nuevo Documento" del menú "Archivo".



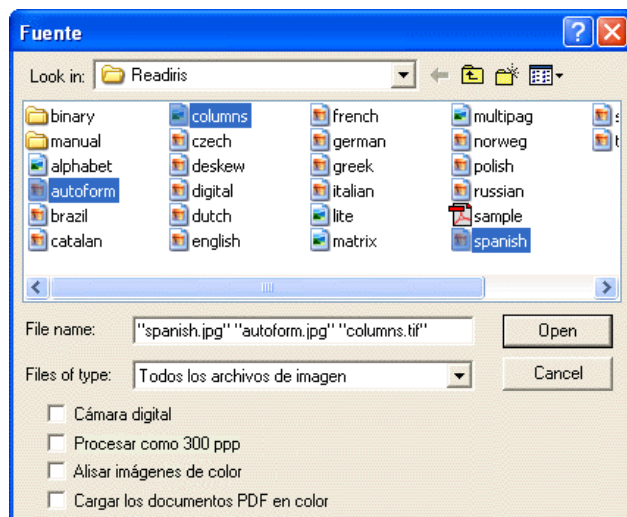
Pero existe un sistema más eficaz para reconocer varias páginas sin tener que escanearlas ni realizar su OCR una por una: ¡se trata del procesamiento directo de **documentos con múltiples páginas**!

Para escanear de una sola vez un documento de varias páginas, deberá activar el alimentador de documentos del escáner, mediante la opción "ADF" (alimentador automático de documentos) bajo el botón "Escáner".



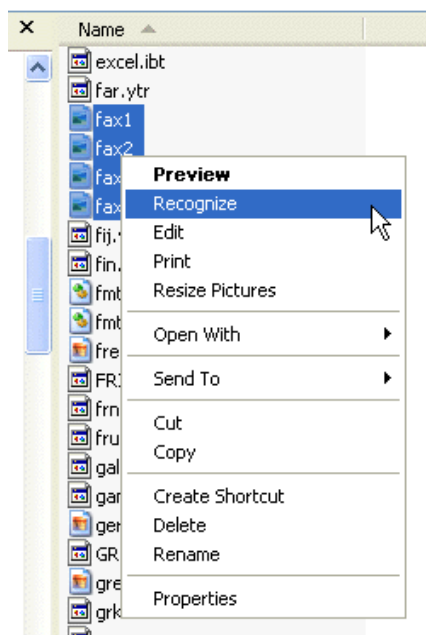
Sitúe todas las páginas en el alimentador automático de documentos y empiece a escanear: se escanearán de la primera a la última página.

También puede *abrir* múltiples imágenes previamente escaneadas. Para ello, seleccione primero una de ellas y, mientras mantiene pulsada la tecla Ctrl, vaya escogiendo las restantes. Para seleccionar un rango continuo de imágenes, seleccione la primera de ellas y, mientras mantiene pulsada la tecla Mayús, escoja la última.





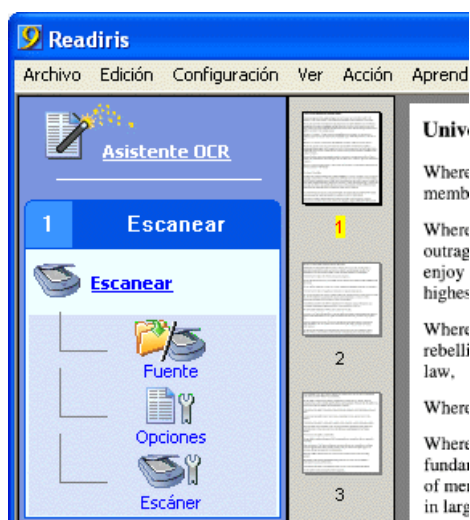
Puede hacer lo mismo desde el Explorador de Windows: seleccione varios archivos de imágenes, haga clic con el botón derecho del ratón y seleccione la opción "Reconocer" del menú contextual. Si repite esta misma operación, todas las imágenes que envíe a Readiris se añadirán al documento en uso hasta que seleccione el mando "Nuevo Documento".



Puede, incluso, abrir imágenes previamente escaneadas *arrastrándolas* desde el Explorador de Windows hasta el área de trabajo de Readiris. También aquí ocurre como citamos anteriormente: puede repetir esta operación cuanto quiera, y todas las imágenes que envíe a Readiris se añadirán al documento en uso hasta que seleccione el mando "Nuevo Documento".

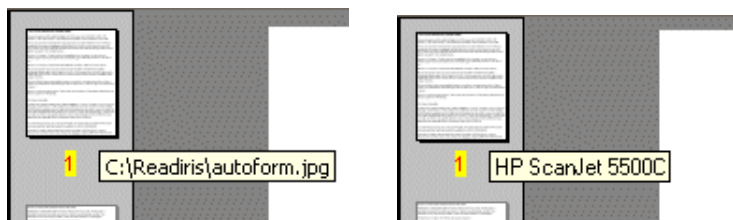
Readiris ordena las imágenes automáticamente - la imagen 001.tif precede a la 002.tif que precede a la 003.tif, etc.

La **barra de herramientas de páginas** situada en la parte izquierda de la imagen - visualizada al procesar las páginas - representa las diferentes páginas del documento; desde ella se accede a los diferentes mandos de páginas (con el botón derecho).

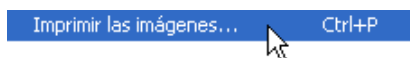


La página activa en un determinado momento queda resaltada en la barra de herramientas de página, y su nombre aparece en la barra de título de Readiris.

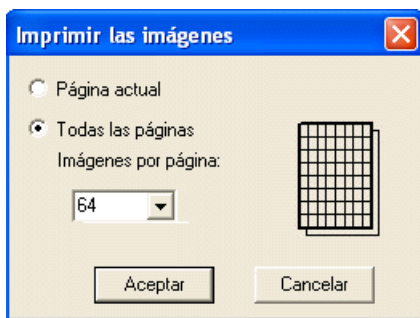
La barra de herramientas de página lleva asociada una etiqueta emergente, que le permitirá conocer la imagen cargada en la memoria con sólo pasar el puntero del ratón por encima de una miniatura de página. (Si se ha abierto una imagen compuesta de múltiples páginas, naturalmente sólo habrá un archivo para todas las imágenes.) Durante el proceso de *escaneo* de documentos con múltiples páginas, la etiqueta emergente tan sólo menciona el modelo de escáner que está empleándose.



Puede rápidamente **imprimir** las **imágenes** numerizadas con el mando "Imprimir las Imágenes" bago el menú "Archivo" si usted necesita un resumen de su documento.

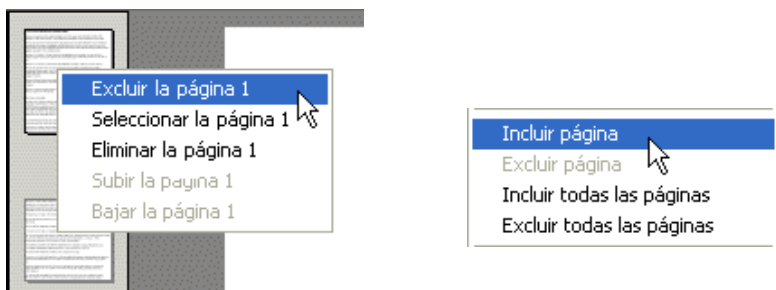


Puede imprimir la página actual o todas las páginas. Selecciones el número de páginas o miniaturas ("thumbnails") que quiera imprimir.

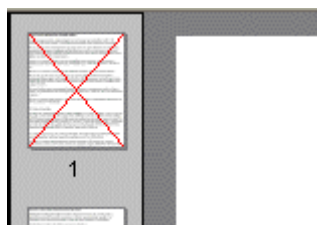


Pero tampoco tiene por qué imprimir todas las páginas: la barra de herramientas de páginas (y el menú "Edición") le permite excluir páginas (temporalmente). Haga clic con el botón derecho en las miniaturas que quiera y seleccione la opción "Excluir la Página" desde el menú "Contexto" o visualice una página y seleccione el mando "Excluir la Página" desde el menú "Edición" para excluirla del proceso de impresión (y reconocimiento). Seleccione la opción "Incluir la

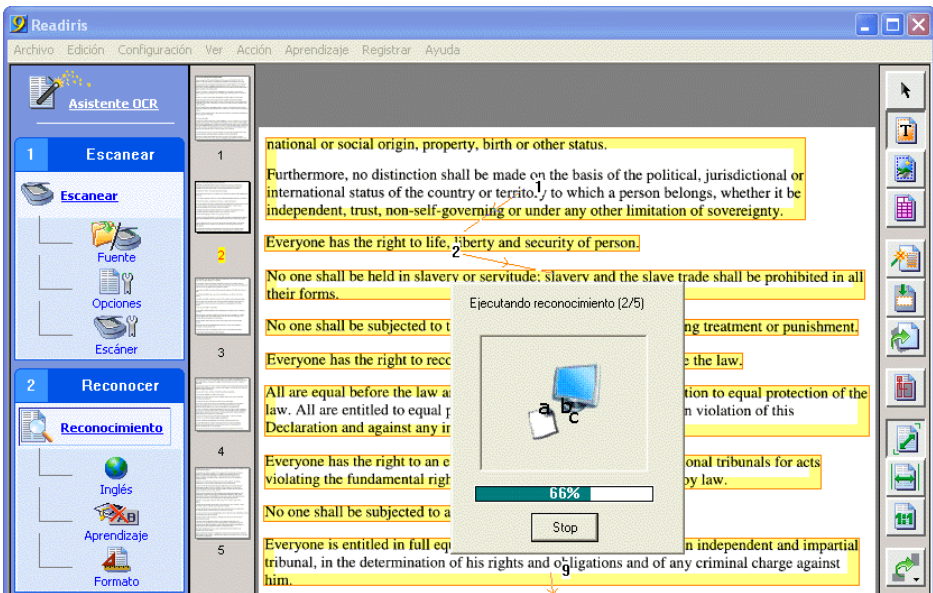
Página" para volver a incluirla. Para conseguir una mayor flexibilidad, el menú "Edición" ofrece opciones equivalentes que se aplican a *todas* las páginas.



Las miniaturas de las páginas excluidas serán tachadas. Tenga en cuenta que siempre funciona la impresión de la página actual, incluso si esta opción está “desactivada” en ese momento.



Abra la imagen de prueba MULTIPAGE.TIF e inicie el reconocimiento. Las páginas van apareciendo unas tras otras. En la barra de título de Readiris se indica el número de la página activa.

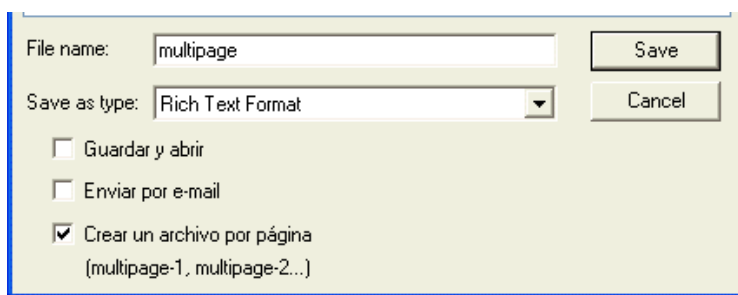


Si el aprendizaje interactivo está en funcionamiento, éste irá reconociendo y aprendiendo página por página. Para la primera de ellas se iniciará un diccionario “nuevo”, y en las sucesivas irá “completándose” este último.

A partir del momento en que pulse el botón "Terminar", el sistema irá tomando decisiones sin que el usuario tenga que validarlas. En otras palabras, se interrumpe el aprendizaje interactivo para *todas* las páginas y el reconocimiento del documento continuará en modo automático.

El resultado del reconocimiento efectuado en documentos con varias páginas queda guardado en un único archivo de salida. (Si se envía el resultado del reconocimiento a una aplicación de salida, se crearán varias páginas dentro de un único documento.)

Al menos, ese es el caso cuando la opción "Crear Un Archivo por Página" esté desactivada al guardar el documento reconocido. Esta opción tratará cada página de un documento multipágina como si estuvieran guardadas en archivos diferentes. Si el usuario pone al archivo el nombre texto.doc, los archivos serán denominados texto-1.doc, texto-2.doc, etc.



EDICIÓN DE DOCUMENTOS CON MÚLTIPLES PÁGINAS

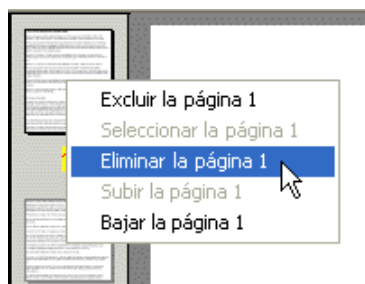
El usuario puede editar documentos con múltiples páginas para corregir, principalmente, errores ocurridos en la fase de escaneo. Podrá eliminar páginas del documento o moverlas a otra posición.

En primer lugar, veamos cómo desplazarnos. Para *ir a una página*, haga clic en su icono de la barra de herramientas de páginas, o bien lleve el puntero a la miniatura, haga clic con el botón derecho del ratón y seleccione la opción "Seleccionar Página" en el menú "Contexto". Para ir a la página anterior, utilice la tecla RePág; para avanzar hasta la página siguiente emplee AvPág. Pulse Home para ir a la primera página, y End para ir a la última página. O utilice las funciones correspondientes en el menú "Ver".

Primera página	Home
Página anterior	PageUp
Página siguiente	PageDown
Última página	End



Y ahora, ocupémonos de la edición de documentos. Para *eliminar una página* de un documento, mantenga el cursor sobre su miniatura, selecciónela con el botón derecho del ratón y escoja el mando "Eliminar la Página". Y le recordamos que podrá excluir páginas temporalmente, pero no borrarlas, del proceso de reconocimiento (y de impresión de imágenes): la barra de herramientas de página (y el menú "Edición") incluyen los mandos necesarias.



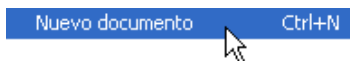
Para *subir una página* dentro de un documento, utilice el mando "Subir la Página". Si, por el contrario, desea *bajarla*, seleccione el mando "Bajar la Página".

Para *mover una página* a una ubicación completamente distinta en el documento, arrastre su icono a esa nueva ubicación.



ABRIR UN NUEVO DOCUMENTO

Para cerrar el documento activo, puede utilizar el mando "Nuevo Documento" del menú "Archivo".



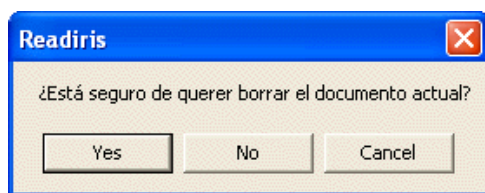
Esta función "limpia completamente la pizarra". Cualquier documento cargado en memoria - ya tenga una o múltiples páginas - se eliminará completamente. A partir de entonces, podrá crear uno nuevo.

Sin embargo, también puede crear un nuevo documento directamente desde el documento activo. Mientras no haya ejecutado el OCR, el sistema entiende que usted desea añadir páginas al documento. Por ejemplo, puede escanear todas las páginas que se encuentren en el alimentador del escáner, colocar más páginas en el alimentador y continuar con el procedimiento. Todas las páginas escaneadas compondrán un único documento. O bien, podrá escanear varias páginas y añadir algunos archivos gráficos, como por ejemplo, faxes. Estas páginas siguen formando



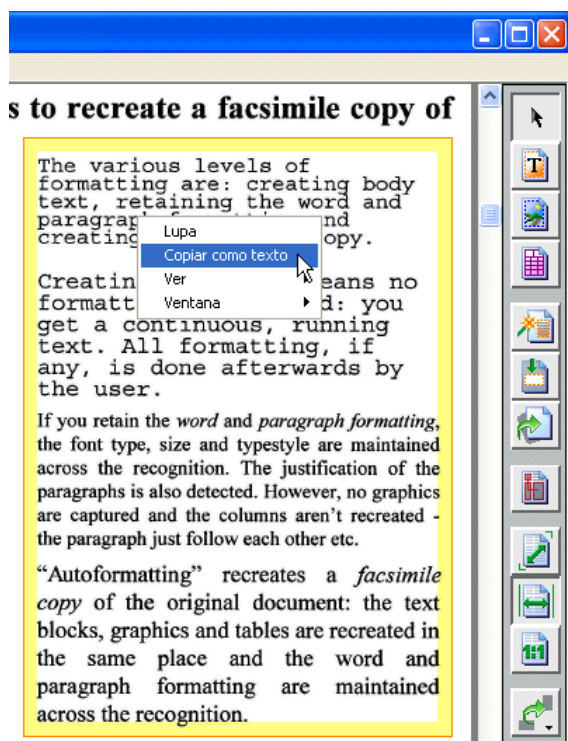
un solo documento; todo lo que tiene que hacer es cambiar el origen de la imagen según vaya siendo necesario. Para ello utilice el botón "Fuente".

Una vez realizado el OCR, si reinicia el proceso de escaneo (o la carga de imágenes), empezará un nuevo documento o completará el que esté en ese momento en uso.



RECONOCIMIENTO DE ZONAS DE TEXTO

Una vez que sabemos cómo reconocer páginas y procesar documentos con múltiples páginas, tenemos que preguntarnos si se puede reconocer menos de una página con la misma facilidad. La respuesta es sí. Para ello, haga clic con el botón derecho del ratón y elija la opción "Copiar como Texto" del menú "Contexto": la ventana del texto bajo el cursor del ratón será reconocida y enviada al portapapeles.



Se aplicará la configuración del sistema actual - idioma, tipo de fuente, etc. El resultado del OCR se colocará en el portapapeles como texto "en corpo", no formateado.



ORGANIZACIÓN DEL TEXTO RESULTANTE

Guardar o exportar el texto requiere algo más que indicar simplemente un método de exportación o un nombre para el archivo resultante. También debe seleccionarse un formato de archivo y determinarse la apariencia del texto reconocido. Es decir, tiene que elegir lo que quiere hacer con el texto antes de iniciar el programa correspondiente.

Algunas opciones del botón "Formato" le permitirán determinar la apariencia del texto resultante.

El **flujo de texto** del documento resultante es determinado directamente por la opción "Fusionar Líneas en los Párrafos".

Opciones

- ☒ Fusionar líneas en los párrafos
- ☒ Incluir gráficos

Mantenga activada esta opción para que Readiris detecte los párrafos. De esta forma, Readiris aplicará el **ajuste de línea** típico de los tratamientos de texto; de lo contrario se añade un retorno de carro (código CR o EOL) tras cada línea y las palabras separadas por guiones quedan tal cual. Por defecto, la detección de párrafos está activada.

Para explicar esto, he aquí un ejemplo. Si las primeras tres líneas de una columna son "El nuevo presi-", "dente saludó desde el balcón." y "Su esposa vino a acompañarlo.", la detección de párrafos le ofrece el siguiente resultado: "El nuevo **presidente** saludó desde el balcón. **Su** esposa vino a acompañarlo." Las partes separadas por el guión de la palabra "presidente" fueron unidas y se añadió un espacio al final de la frase, creando así un flujo natural de texto.

Si la detección de párrafos *no* estuviese activada, se habría mantenido el diseño original con un retorno de carro añadido al final de cada línea.

Esta opción *no* está disponible cuando el formato PDF está seleccionado: siempre se almacenan los archivos PDF Adobe Acrobat línea por línea.

(El botón "Formato" contiene algunas opciones que aún no hemos explicado; lo haremos pronto.)

CONFIGURACIÓN DEL ESCÁNER

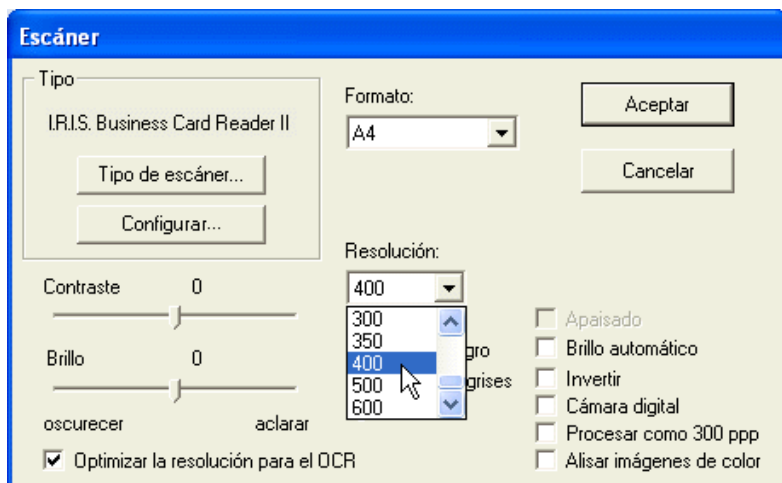
Ahora debe configurar su escáner. Se supone que el escáner y los controladores necesarios han sido instalados correctamente.

Si la licencia de su *software* Readiris va unida a un modelo de escáner, probablemente este paso resulte innecesario, ya que su escáner puede estar configurado para Readiris.

Haga clic en el botón "Escáner" de la barra principal de herramientas.



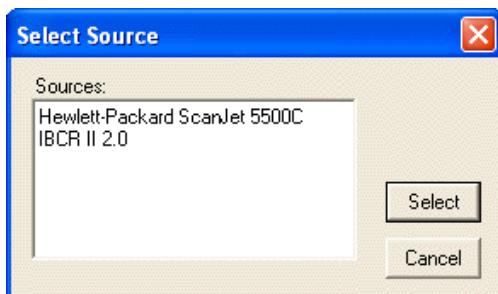
Haga clic en el botón "Tipo de Escáner" para determinar el **modelo de escáner**.





Si selecciona la opción "<Image>" (Imagen) como "escáner" - las imágenes ya escaneadas funcionan como origen de las imágenes en todo momento - no tiene que seleccionar el disco como origen de imagen con el botón "Fuente" de la barra principal de herramientas.

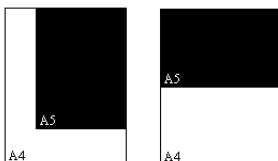
El botón "Configurar" sólo está disponible si su escáner lo permite. Para acceder a algunos parámetros avanzados de escaneado con escaners Twain, haga clic en el botón "Configurar" y seleccione el origen Twain. (También puede utilizar la opción "Seleccionar Fuente" del menú "Archivo".)



Una vez seleccionado el escáner, la misma ventana le permitirá establecer la resolución de escaneado, el formato y la orientación de la página y el brillo y el contraste. Asimismo, tendrá ocasión de indicar si va a utilizar el alimentador del escáner. Con escaners Twain, todos los parámetros de escaneado a menudo se establecen dentro del interfaz Twain.

Establezca el **brillo** y, si dispone de él, el **contraste**.

Si activa la opción "Apaisado" indicará que la página tiene dicha orientación y no la vertical. En realidad, la orientación de la página se aplica a los formatos reducidos de página: en un escáner plano de formato A4, se pueden escanear páginas de formato A5 (la mitad del tamaño del A4) en orientación vertical y horizontal, pero obviamente sólo puede escanear la superficie completa de una página A4 en una única dirección.



Por último, la opción "Invertir" le permitirá generar **imágenes “invertidas”** en la modalidad de escaneo blanco y negro - puede activar esta opción para procesar páginas enteras en las que aparezca texto blanco sobre fondo negro.

PONGA COLORES EN SUS TEXTOS DIGITALIZADOS!

Readiris suporta sin ningún problema imágenes en blanco y negro, en escala de grises y en color. Escoja, pues, la **modalidad de color** que mejor se adapte a sus necesidades. Si va a incluir en los documentos dibujos, escanee en blanco y negro, si desea incorporar fotografías en blanco y negro, escanee en escala de grises y para insertar fotografías en color, escanee en color.

Pero ¿cuál es la razón para reducir la profundidad de las imágenes en la fase de escaneo? Ni que decir tiene que las imágenes en escala de grises y en color se escanean con más lentitud y requieren más memoria RAM que las imágenes en blanco y negro.

Escanear en escala de grises o en color no es sólo útil para guardar imágenes con suficiente calidad, ¡a veces, también puede ser útil - o incluso necesario - para obtener buenos resultados con el OCR! Cuando el texto está impreso sobre un fondo coloreado, el hecho de escanear en color puede generar esos matices tonales de los que carecen las imágenes en blanco y negro. Cuando el contraste entre el texto y el fondo es reducido, este último puede crear cierto “ruido”; ¡lo que dificultará enormemente el reconocimiento, o lo hará imposible!

Piense, por ejemplo, en un texto negro impreso sobre fondo oscuro. Si escanea un texto así en blanco y negro, por mucho que intente ajustar el brillo del escáner, no podrá *deshacerse* del color de fondo sin perder también la información del texto...



MASAYOSHI SON, 42, president and CEO, is the master Net empire builder. His conglomerate holds stakes in 300 Internet companies in the U.S., Japan, Europe, and other Asian countries. Today, Softbank manages about \$4 billion in venture capital funds for global investments.

YASUMITSU SHIGETA, 35, has invested in more than 70 Web or mobile Net-based ventures in Japan and the U.S., including Tumblweed Communications and Phone.com. Shigeta is also developing new businesses that take advantage of the growth of the Internet and mobile communications.

MASAYOSHI SON, 42, president and CEO, is the master Net empire builder. His conglomerate holds stakes in 300 Internet companies in the U.S., Japan, Europe, and other Asian countries. Today, Softbank manages about \$4 billion in venture capital funds for global investments.

YASUMITSU SHIGETA, 35, has invested in more than 70 Web or mobile Net-based ventures in Japan and the U.S., including Tumblweed Communications and Phone.com. Shigeta is also developing new businesses that take advantage of the growth of the Internet and mobile communications.

Readiris genera una versión en blanco y negro de cada imagen en escala de grises o en color. Sus rutinas inteligentes resuelven incluso casos muy problemáticos. Así es como se digitalizan las imágenes “difíciles”.

MASAYOSHI SON, 42, president and CEO, is the master Net empire builder. His conglomerate holds stakes in 300 Internet companies in the U.S., Japan, Europe, and other Asian countries. Today, Softbank manages about \$4 billion in venture capital funds for global investments.

YASUMITSU SHIGETA, 35, has invested in more than 70 Web or mobile Net-based ventures in Japan and the U.S., including Tumblweed Communications and Phone.com. Shigeta is also developing new businesses that take advantage of the growth of the Internet and mobile communications.

Para visualizar una imagen escaneada en blanco y negro, deshabilite la opción "Visualizar Documento en Color" del menú "Ver".

✓ Visualizar documento en color Ctrl+O



Si esta opción está activada, usted no verá ninguna imagen en blanco y negro en la pantalla de su ordenador, incluso si se escanean imágenes en dos niveles. Esto se debe a que la opción "Visualización Optimizada" del menú "Ver" optimiza las imágenes para una perfecta legibilidad en pantalla.



La técnica especializada de alta resolución transforma las imágenes en blanco y negro en imágenes en escala de grises.

Reading dot matrix documents

You can read dot matrix document without changing the font mode. The software detects whether "normal" text or dot matrix printouts are being read.

Far out in the uncharted backwaters of the unfashionable end of the Western Spiral arm of the Galaxy lies a small unregarded yellow sun. Orbiting this at a distance of roughly ninety-two million miles is an utterly insignificant little blue green

Reading dot matrix documents

You can read dot matrix document without changing the font mode. The software detects whether "normal" text or dot matrix printouts are being read.

Far out in the uncharted backwaters of the unfashionable end of the Western Spiral arm of the Galaxy lies a small unregarded yellow sun. Orbiting this at a distance of roughly ninety-two million miles is an utterly insignificant little blue green

Se suavizan y difuminan las imágenes en escala de grises y en color.

A word about OCR

The aim of OCR is to automatically enter printed text documents in a very effective and low cost way. Although the first research and development on Optical Character Recognition (OCR) began more than 30 years ago, this technology is still unknown by most of the people who could use it for their document entry applications.



A word about OCR

The aim of OCR is to automatically enter printed text documents in a very effective and low cost way. Although the first research and development on Optical Character Recognition (OCR) began more than 30 years ago, this technology is still unknown by most of the people who could use it for their document entry applications.

Como consecuencia, no será necesario ampliarlas, incluso en portátiles con pantallas de cristal líquido u ordenadores de mesa con pantallas de 14" de baja resolución. La visualización de alta calidad está activada por defecto, pero puede ser innecesaria en ordenadores con pantallas de alta resolución.

APARATOS DISTINTOS, RESOLUCIÓN DISTINTA

Sea como sea su modo de numerización, utilice una **resolución** de escaneo de 300 ppp para aplicaciones normales. Cuando el documento que vaya a escanear tenga un tipo de letra pequeño (menos de 10 puntos), o sea de mala calidad, utilice mejor una resolución más alta, de 400 ppp.

Readiris lee **tamaños de punto** entre 6 y 72 puntos (0,21 y 2,54 cm.).

6 puntos

72 puntos

Readiris reconoce asimismo “**letras descendentes**” (“drop letters”), mayúsculas de gran tamaño que cubren varias líneas. (Desde luego, éstas no pueden ser superiores a 72 puntos.)

Readiris reconoce letras descendentes, mayúsculas de gran tamaño que cubren varias líneas.

Dado que el OCR óptimo requiere una resolución de entre 300 y 400 dpi, Readiris le avisa cuando escanea imágenes con una resolución inferior a 200 dpi o superior a 800 dpi. Sin embargo, Readiris puede corregir los imágenes escaneadas con demasiados detalles. Para llevar a cabo la corrección, active la opción "Optimizar la Resolución para el OCR" en la configuración del escaneado. Siempre que la resolución de la imagen de los escaneados supere los 600 dpi, la resolución se reduce para el proceso del OCR.

☒ Optimizar la resolución para el OCR

Existen otras maneras de evitar este aviso: cuando lee **faxes** (que presentan una resolución de 100 o 200 dpi), cuando genera imágenes con una cámara digital (cuya resolución se desconoce) y cuando abre imágenes en las que el encabezamiento del archivo contiene una resolución incorrecta. Para procesar este tipo de imágenes sin ningún problema, active la opción "Procesar como 300 ppp". Este parámetro sirve para digitalizar directamente y abrir imágenes previamente escaneadas.

☐ Invertir
☐ Cámara digital
☒ Procesar como 300 ppp
☐ Alisar imagen de color

☐ Cámara digital
☒ Procesar como 300 ppp
☐ Alisar imágenes de color
☐ Cargar los documentos PDF en color

Cuando digitaliza imágenes con una **cámara digital** en lugar de un escáner, es imprescindible activar una opción específica. (Este parámetro sirve otra vez para digitalizar directamente y abrir imágenes previamente escaneadas.)



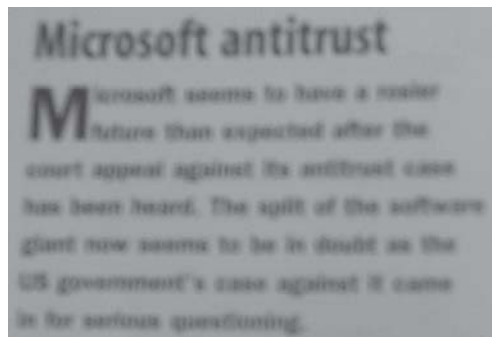
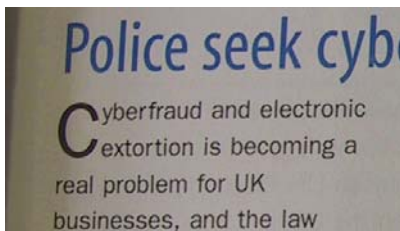
<input type="checkbox"/>	Invertir
<input checked="" type="checkbox"/>	Cámara digital
<input checked="" type="checkbox"/>	Processar como 300 ppp
<input type="checkbox"/>	Alisar imagen de color

<input checked="" type="checkbox"/>	Cámara digital
<input checked="" type="checkbox"/>	Processar como 300 ppp
<input type="checkbox"/>	Alisar imágenes de color
<input type="checkbox"/>	Cargar los documentos PDF en color

Así, optimiza la imagen antes de reconocerla. Existen varias dificultades en el momento de digitalizar con cámaras digitales: generan imágenes con baja resolución - incluso cuando dispone la cámara muy cerca al documento - y en todo caso no se conoce la resolución de la imagen.

Tendrá que tener en cuenta algunos parámetros delicados a la hora de digitalizar correctamente imágenes escaneadas con una cámara digital.

Antes de todo, seleccione una resolución de imagen la más alta posible. Genere por ejemplo imágenes con una talla de 2.048 x 1.536 cuando se reconocen también las imágenes 1.024 x 768 y 640 x 480. Luego, active el modo “macro” de su cámara para los primeros planos - siempre es así cuando fotografía documentos. (Este modo ha sido diseñado para capturar flores, insectos, etc.) En el caso contrario las imágenes serían imprecisas e ilegibles.

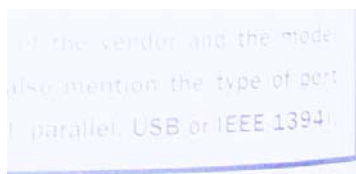


Tendrá que limitarse a pequeñas compresiones o a ninguna: una compresión importante reduce la nitidez del texto digitalizado. Puede agrandecer para poder

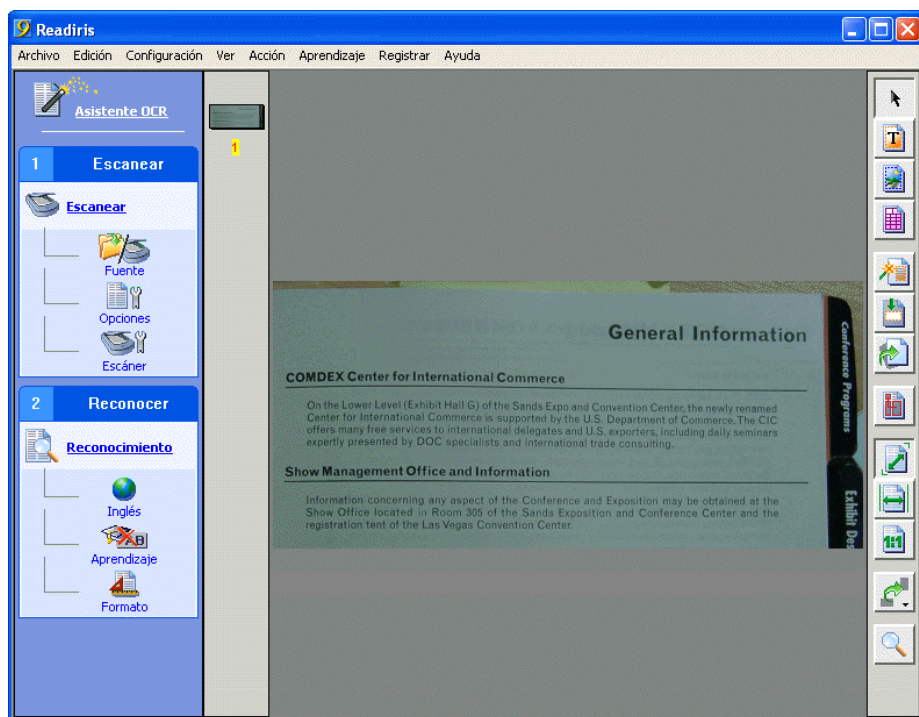
recortar sus documentos - algunas cámaras vienen incluidas con un programa para restablecer la fotografía original ("*photo stitching*"), pero no es útil para capturar un documento.

Mantenga la cámara directamente encima del documento para evitar capturar el documento bajo un cierto ángulo. Sin embargo, evite que la sombra del aparato o de su mano sea proyectada en el documento. Genere imágenes estables. Piense también en colocar su cámara en un trípode cuando sea necesario.

Desactive el flash cuando fotografía papel brillante, o sino puede ser que la imagen sea demasiado pálida. Generalmente, mejor vale adaptar el brillo y contraste al entorno - luz del día, luz de una lámpara, luz de neón, etc. (Algunas cámaras puede ser calibradas filmando un documento blanco.)



Para dar un ejemplo, abre la imagen DIGITAL.JPG en la carpeta Readiris y ejecute el reconocimiento.



CÓMO GUARDAR LAS PROPIEDADES PREDETERMINADAS

Establezca correctamente todos los parámetros de escaneado y haga clic en la opción "Guardar la Configuración por Defecto" del menú "Archivo" para guardar las propiedades actuales como propiedades predeterminadas para su uso posterior.



Guardar la configuración por defecto

Los archivos de configuración no sólo contienen la **configuración** del escáner, sino que también determinan si desea utilizar el aprendizaje interactivo, el idioma de los documentos, qué modo de salida se utilizará - por ejemplo, enviar el texto a la aplicación WordPad - etc. En pocas palabras, *todas* las propiedades de funcionamiento de Readiris se almacenan en los archivos de configuración.

CÓMO GUARDAR PROPIEDADES ESPECÍFICAS

Obviamente las propiedades predeterminadas se utilizarán cada vez que se inicie el programa, pero puede guardar propiedades específicas para no tener que redefinir estos parámetros de funcionamiento. Las opciones "Guardar Configuración" y "Cargar Configuración" del menú "Archivo" se encargan de ello.



Cargar configuración...

Guardar configuración...

Guardar la configuración por defecto

He aquí un ejemplo: si periódicamente debe reconocer documentos en español con un diseño específico, le recomendamos que cree un archivo de propiedades para dicho tipo de documento. A continuación seleccione "Español" como idioma del documento, cargue una división en zonas concreta para no tener que trazar las mismas ventanas a cada página, desactive el aprendizaje, pero active un diccionario de fuentes en modo "de lectura" porque algunas tipos de fuentes se usan de forma sistemática, etc.

Si no está seguro de las propiedades actuales, no necesita pasar revista a todos los menús y opciones para descubrirlas. Utilice la opción "Información" del menú "Archivo" para obtener una sinopsis de las mismas.



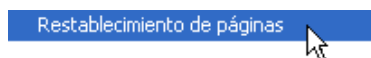
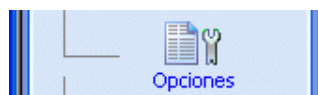
Configuración			
Escáner		<input type="button" value="Aceptar"/>	
Modelo	<Twain>		
Resolución	300 ppp		
Formato	A4 300		
Brillo	0		
Contraste	0		
Modo	Blanco-negro		
Apaisado	Inactivo		
		Texto	
		Formato	Microsoft Word 97 / 2000 / 2002 / 2003
		Párrafo	Activo
		Formato	Volver a crear el documento original
Documento			
Fuente	Automático	Idioma	Español
Página			
Resolución	300 ppp		

NUMERIZAR DOCUMENTOS

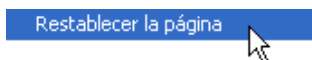
Ahora que el escáner está correctamente configurado, vamos a empezar a escanear documentos. Hay algunos aspectos de los que debe estar informado.

En primer lugar, tenga cuidado con la inclinación de las líneas. Aunque el análisis de página y el reconocimiento admiten cierto grado de inclinación, la división en ventanas y el OCR de una página pueden verse dificultados si la inclinación de las líneas es excesiva. Una pequeña inclinación (menos de 0,5°) no tiene ninguna importancia: la fidelidad del OCR no se ve afectada.

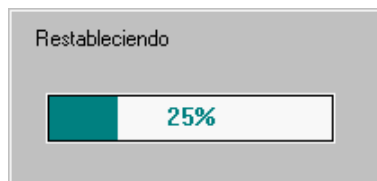
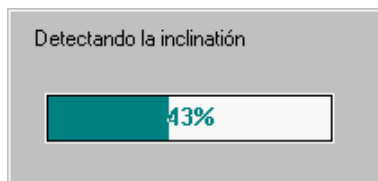
El mando "Restablecimiento de Páginas" del botón "Opciones" (y del menú "Configuración") determina si las páginas escaneadas con alguna inclinación van a **restablecerse**, se arreglarán automáticamente. (Las pequeñas inclinaciones no se tendrán en cuenta.) Esta opción se encuentra desactivada por defecto.



Si ha olvidado activar esta opción, utilice el botón "Restablecer la Página" de la barra de herramientas de imágenes (o la opción del mismo nombre del menú "Acción") para “rectificar” las páginas escaneadas con alguna desviación.



Este restablecimiento de líneas tarda algunos segundos, dado que, en primer lugar, se analiza la imagen para detectar el posible ángulo de inclinación. Seguidamente se recompone la imagen en color o en escala de grises (así como su versión en blanco y negro) y, por último, vuelve a ejecutarse el análisis de página.



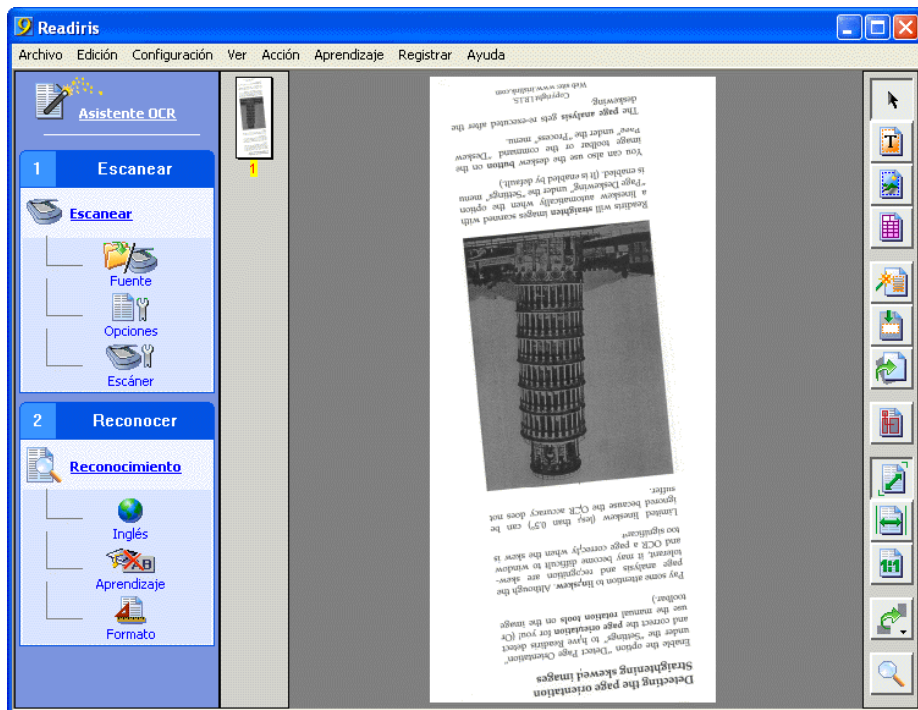
Cada vez que necesite ajustar la orientación de la página, puede usar las herramientas de **rotación** de la barra de herramientas de imágenes. (Los mandos correspondientes se encuentran en el menú "Ver".) Existen tres direcciones de rotación: a la izquierda, a la derecha e inversión. La rotación tarda algunos segundos en actualizar la imagen en sí y no sólo lo que se muestra en pantalla.



Sin embargo, Readiris puede corregir por usted las páginas mal orientadas. Active la opción "Detectar la Orientación de las Páginas" bajo el botón "Opciones" (o en el menú "Configuración") y Readiris orientará la página cuando sea necesario.



Si desea comprobar cómo funciona esta opción, utilice la imagen DESKEW.JPG que se encuentra en la carpeta Readiris. Desactive los opciones "Restablecimiento de Páginas" y "Detectar Orientación de las Páginas" antes de abrir la imagen y deje que Readiris restablezca la Torre de Pisa... tal como la conocemos.



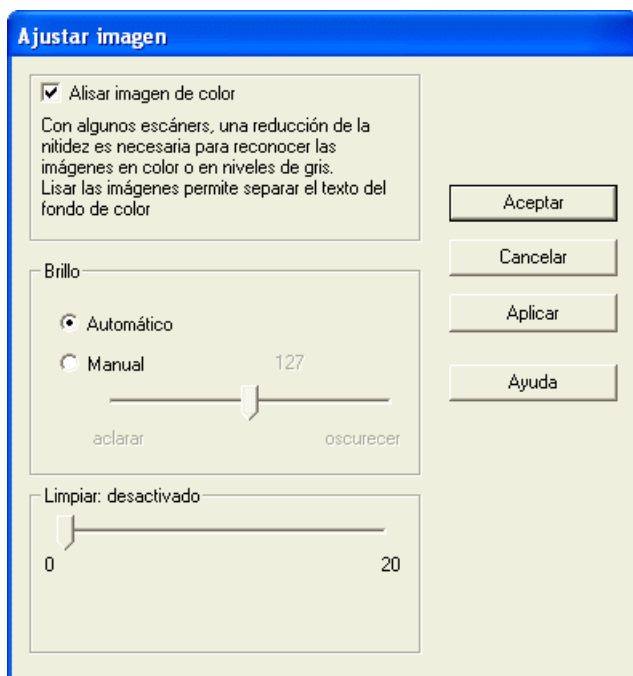
AJUSTE DE LAS IMÁGENES ESCANEADAS

Como ya hemos indicado, las imágenes en color o escala de grises se transforman automáticamente en imágenes en blanco y negro mediante avanzadas rutinas inteligentes. En caso de que siguiera siendo necesario, el usuario puede afinar aún más la imagen, con vistas a un mejor OCR. Seleccione para ello la opción "Ajustar Imagen" del menú "Acción".



Ajustar imagen... Ctrl+J

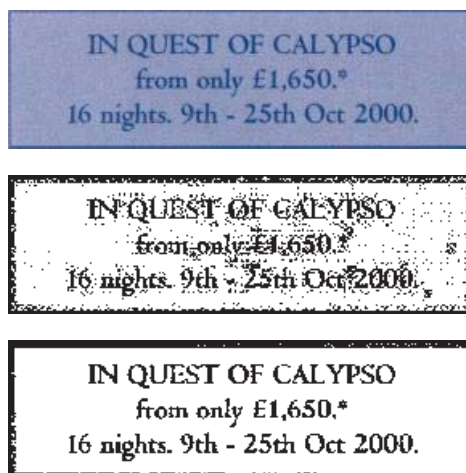
Cuando seleccione esta opción, se visualizará automáticamente la versión en blanco y negro de la imagen - es como si desactivara la opción "Visualizar Documento en Color". Hay algunos conceptos de cierta complejidad, que pasamos a desarrollar con detalle seguidamente.



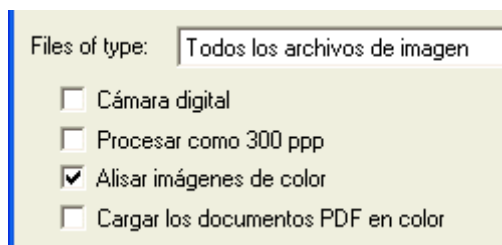
La opción "Alisar Imagen de Color" da a las imágenes en color o escala de grises un aspecto más homogéneo y para ello "rasea", es decir, suaviza las

diferencias relativas de intensidad, con lo que el contraste queda más afinado entre el primer plano - texto - y el fondo - un color, ilustraciones, etc.

Puede que este dispositivo de **preprocesamiento** le parezca muy técnico y difícil de entender, pero es realmente funcional, ya que con algunos modelos de escáner se necesita esta reducción de nitidez para reconocer imágenes en color y escala de grises. A veces, el único modo de separar el texto del fondo de color consiste en suavizar las diferencias. Abajo tiene un ejemplo de imagen que es ilegible si no se suaviza la imagen.



Se puede también suavizar las imágenes previamente escaneadas cuando se cargan en la memoria.



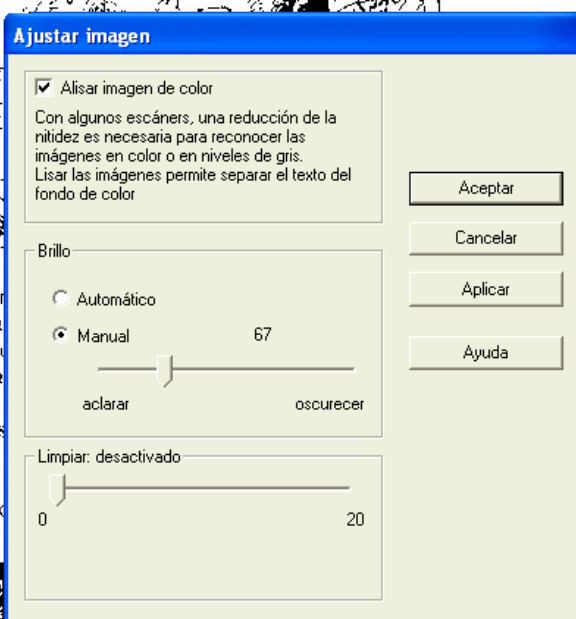


Ocupémonos ahora del brillo. Cuando hablamos de “brillo”, nos referimos en concreto al umbral de blanco y negro. Al seleccionar la opción "Automático" el ajuste entre ambos colores se determina automáticamente. Aplique un umbral diferente, cuando sea necesario, oscureciendo o aclarando la imagen en blanco y negro. Al oscurecer la imagen, hay una mayor cantidad de píxeles negros en la versión en blanco y negro de la imagen. Por el contrario, al aclarar la imagen, una mayor cantidad de píxeles blancos se instauran en la versión en blanco y negro.

Observe que no se ejecuta ningún ajuste de imagen hasta que no haga clic en el botón "Aplicar". Al pulsar "Aceptar" se ejecutará el ajuste y se cerrará la ventana. Éste es un ejemplo en el que se ha aclarado drásticamente una imagen en blanco y negro. ¡Naturalmente, sin pretender en ningún momento que el OCR sea de calidad!

El documento es leído por un
imagen. En esta etapa, la ima
pixels, sobre un fondo blanco
pixels: tiene que reconocer la

Utiliza bases de datos lingüís
así una potente herramienta
capacidad de aprendizaje, se
diferentes tipos de documento
de caracteres.



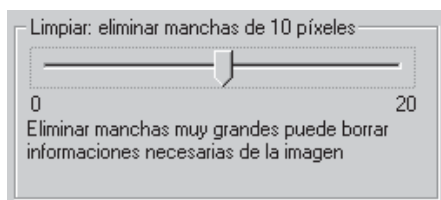
Las dos primeras opciones son sólo válidas para las imágenes en color o en escala de grises. La última de ellas, "Limpiar", concierne exclusivamente a las imágenes en blanco y negro. Por "limpieza" entendemos la eliminación de esa "suciedad" (motas) que ocasionan algunos pixeles innecesarios.

**If computers can't
adapt easily, then
maybe the people
using them can.**

**If computers can't
adapt easily, then
maybe the people
using them can.**



Asegúrese de no borrar motas demasiado grandes, pues si no podría eliminar también los puntos de las "i", o partes de letras obtenidas mediante impresoras matriciales, etc.



Para conseguir la mejor calidad de imagen con vistas a un buen OCR, coloque la ventana de ajuste de imagen donde no moleste para apreciar los arreglos que vaya haciendo. Seguidamente, ajuste los parámetros y vaya pulsando "Aplicar" hasta conseguir una imagen clara y nítida.

DEJE QUE EL ASISTENTE TRABAJE POR USTED

Pasemos ahora a capturar documentos. En lugar de configurar uno a uno los parámetros, utilizaremos el **asistente de OCR**. Éste es un sistema muy sencillo para reconocer páginas.

Haga clic en el botón "Asistente OCR" de la barra de herramientas principal (o seleccione la opción "Asistente OCR" del menú "Acción").



El asistente lo ayudará durante el proceso de reconocimiento. Bastará con responder a unas simples preguntas para obtener rápidos resultados.



Observe que el asistente de OCR se ejecuta cada vez que inicia Readiris; para que esto no ocurra, desactive la opción "Activar el Asistente al Iniciar", que aparece en la primera pantalla del asistente. (También puede desactivarlo usando la misma opción del menú "Configuración".)

READIRIS REPRODUCE EL DISEÑO DE SUS DOCUMENTOS

El asistente de OCR realiza el proceso de reconocimiento de forma prácticamente automática, ¡pero el reconocimiento “automático” *no* debe confundirse con el formato automático! “Formato automático” significa que Readiris reproduce una **copia facsímil** del documento escaneado, aplicando el formato de las palabras, párrafos y páginas del documento original.

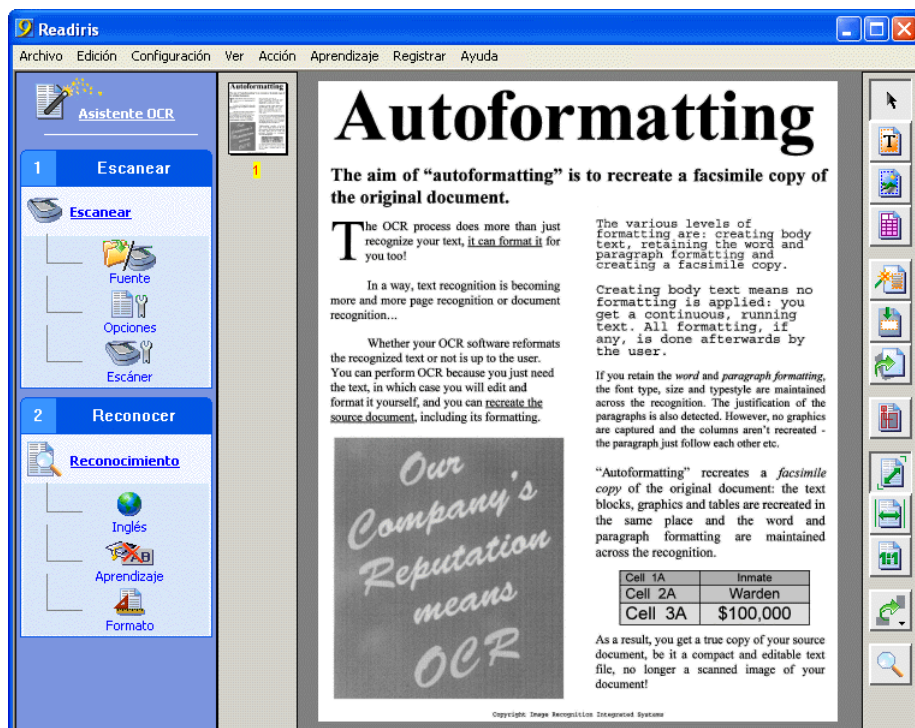


Se utilizan tipos de letra iguales a los del documento original (serif y sans serif, proporcionales y fijos, normales y condensados, superíndice y subíndice). Se mantienen los tamaños y estilos (negrita, cursiva, subrayado, superíndice y subíndice) durante el reconocimiento. Se reproducen las tabulaciones y la justificación (izquierda, centrada, derecha o justificada) de cada bloque de texto. Lo mismo ocurre con las viñetas y la numeración. Todas las direcciones de correo electrónico y de páginas web serán detectadas y se establecerán como hipervínculos en el texto creado. La colocación de columnas, bloques de texto y gráficos será la misma del documento original.

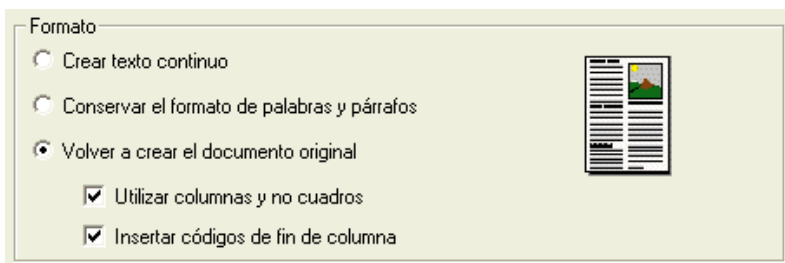
En otras palabras, Readiris le permite archivar una copia fiel de sus documentos, ¡pero ahora en forma de archivo de texto digital, modificable y compacto, en lugar de una imagen escaneada!

Así pues, la clasificación de ventanas sólo se aplica *parcialmente* al realizar el “formato automático”. Es decir, ¡puede incluir y excluir zonas, pero cualquier reclasificación no será tomada en cuenta!

He aquí un ejemplo de cómo funciona. Para familiarizarse con esta función, abra el archivo gráfico AUTOFORM.JPG que se encuentra en la carpeta Readiris.

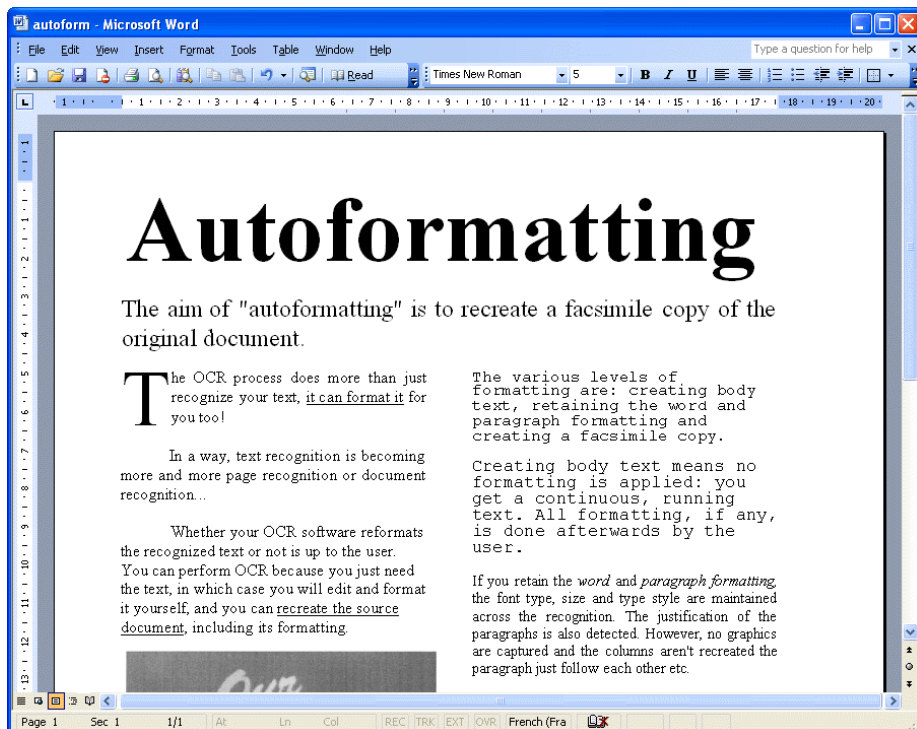


Haga clic en el botón "Formato" de la barra de herramientas principal y envíe el resultado del reconocimiento a Microsoft Word o bien seleccione "RTF" (Rich Text Format) o Word (DOC) como formato de salida. En segundo lugar, seleccione "Volver a Crear el Documento Original" como opción de formato. (La opción "Fusionar Líneas en los Párrafos" está activada por defecto para realizar el ajuste de líneas dentro de los párrafos.)

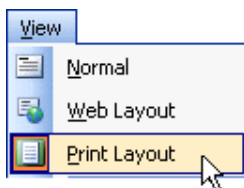


Que la reconstrucción del formato sea posible o no depende del modo de salida. Algunos formatos “pobres” que generan texto “simple”, como Texto (ANSI), Texto formato MS-DOS (ASCII), etc., *no* pueden utilizar los códigos avanzados de formato y, por consiguiente, no pueden ofrecer un formato automático. Por otro lado, el formato PDF Adobe Acrobat se diseñó para copiar el aspecto de los documentos. Así pues, los documentos PDF implican, por naturaleza, el formato automático.

Al abrir el texto reconocido con un tratamiento de textos, el texto tiene esta apariencia, sin intervención *alguna* del usuario.



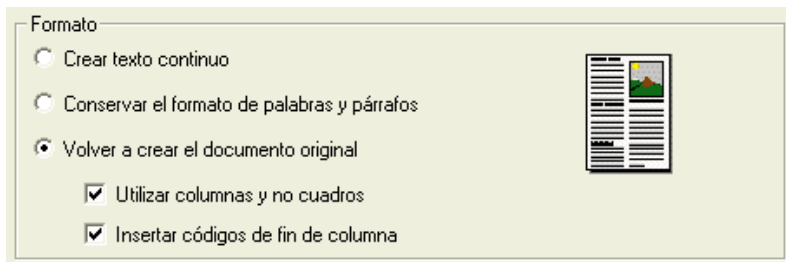
Para ver este efecto correctamente, necesita activar el modo “WYSIWIG” del procesador de textos, denominado generalmente modo “diseño de página”. ¡No obstante, si envía directamente el documento reconocido a Microsoft Word, la vista “diseño de página” o “diseño de impresión” se activa automáticamente!



En pocas palabras, Readiris no sólo reconoce sus textos, sino que puede también formatearlos. Ahora, el OCR ya no es un simple reconocimiento de textos, sino que cada vez se acerca más a un **reconocimiento de páginas** y documentos.

UTILICE COLUMNAS EN LUGAR DE CUADROS

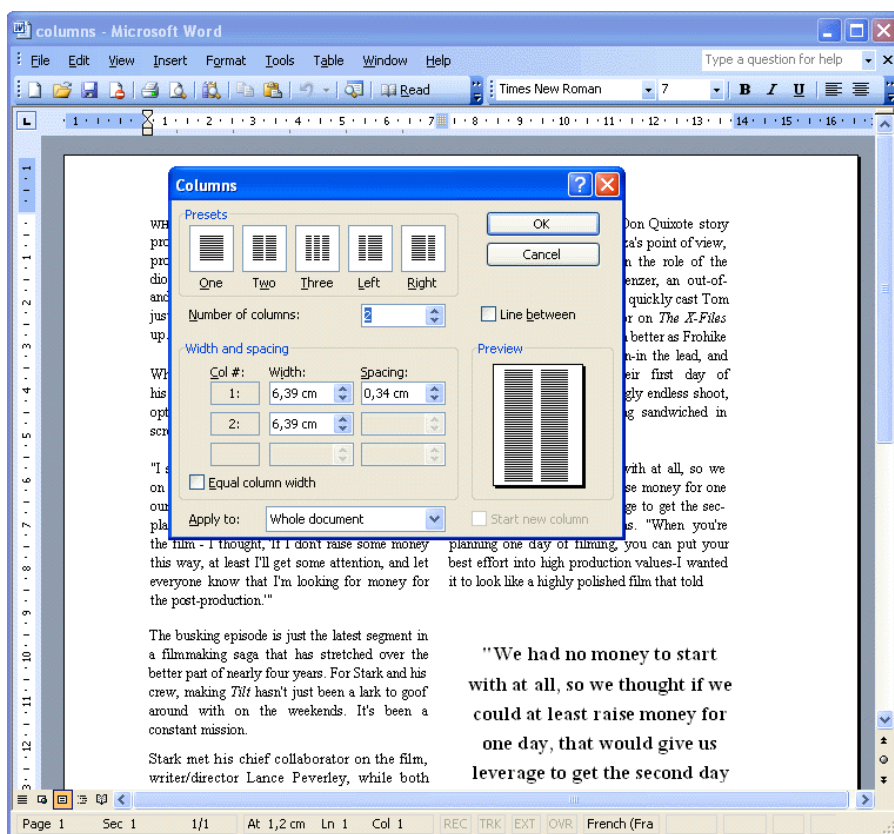
La opción de formato "Utilizar Columnas y No Cuadros" determina *cómo* se realiza el "formato automático": los bloques de texto, las tablas y los gráficos pueden almacenarse en cuadros o en **columnas** que pueden editarse.



Los "cuadros" son contenedores de texto separados que sirven para colocar varios bloques de texto, de gráficos y de tablas en una página. El texto en columnas pasa de forma natural de una columna a la siguiente y es mucho más fácil de editar.

Imaginemos ahora que en el documento escaneado hay columnas reales. Cuando el sistema es incapaz de detectar columnas en el documento original, este modo de formato utiliza igualmente cuadros por defecto!

Si quiere probarlo, puede utilizar la imagen COLUMNS.TIF de la carpeta de Readiris.



La opción "Insertar Códigos de Fin de Columna" mejora la recreación de columnas: dermina si desea o no insertar un código de fin de columna entre ellas.



Con los códigos de fin entre columnas, cualquier texto que se edite, añada o borre permanecerá dentro de su columna; ningún texto fluirá de forma automática a la siguiente columna. Aquellos textos que vayan después de un espacio entre columnas serán colocados al principio de la columna siguiente.

Active esta opción si quiere conservar los fin de columnas en el mismo lugar en el que se hayan detectado en el documento reconocido - independientemente de la edición de texto que se realice después del OCR. En periódicos y revistas, las columnas de una página se corresponden casi siempre a distintos artículos. Por lo tanto, no es buena idea permitir que el texto fluya automáticamente de una columna a la siguiente.

Desactive esta opción cuando haya dividido el texto en columnas: así se asegurará la fluidez natural del texto de una columna a la siguiente.

FORMATO DE TEXTOS, 2ª PARTE

Las otras opciones de diseño son "Crear Texto Continuo" y "Conservar el Formato de Palabras y Párrafos".

Como ilustra el icono de la derecha, crear **texto de cuerpo** significa crear un texto “continuo”, sin formato. El texto será capturado, pero se ignorará totalmente el formato. Utilice esta opción cuando necesite capturar un texto pero no su diseño.

El cuerpo del texto es también lo que usted selecciona cuando reconoce rápidamente una zona del texto haciendo clic con el botón derecho, y seleccionará

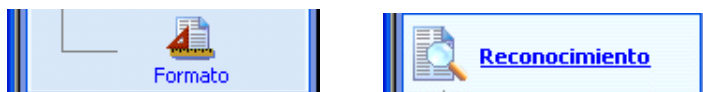
la opción "Copiar como Texto": Una vez realizado el reconocimiento, pegará usted el cuerpo del texto en su aplicación de texto.

La opción "Conservar el Formato de Palabras y Párrafos" representa un camino intermedio: durante el reconocimiento se mantiene el **formato de palabras** - tipo de letra, el tamaño de letra y el estilo. Asimismo, también se mantiene el **formato de párrafo** - tabulaciones y justificación.

Pero no hay que confundir este formato con el formato automático "completo": la primera opción coloca los párrafos uno tras otro y no reproduce columnas, ni copia la posición relativa de las diferentes zonas.

EXPORTAR EL TEXTO VARIAS VECES

En realidad, podrá exportar los resultados del OCR varias veces sin tener que repetir el reconocimiento. Cambie el formato del texto y las opciones de formato con el botón "Formato" y haga de nuevo clic en el botón "Reconocimiento". En este caso, no se ejecutará OCR, a menos que haya usted definido nuevas ventanas o haya modificado las existentes. De lo contrario, Readiris sólo cambiará el formato de los resultados OCR y los guardará en el nuevo formato del texto o los enviará a la aplicación de llegada que usted haya seleccionado.



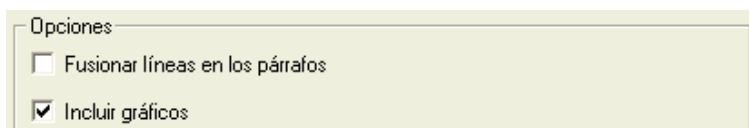
Lo mismo se aplicará para cualquier otro elemento que usted cambie: cuando añada una página a su trabajo OCR, sólo se reconocerá esa página. Si crea una nueva zona de texto en una página, sólo será reconocida esa zona antes de que los resultados se exporten.

Por ejemplo, usted podría reconocer un documento de 10 páginas y guardarlo en un archivo de Word, escanear el resumen de la primera página y enviarlo por correo electrónico a un compañero impaciente, escanear el anexo - una tabla - y guardar todos los resultados en un archivo HTML para colgarlo en la página web de su empresa.



GUARDAR GRÁFICOS POR SEPARADO

En nuestro ejemplo, el gráfico se incluía en el texto reconocido; que esto sea así o no, dependerá de la opción de formato "Incluir Gráficos". Los gráficos que hay en el texto se podrán volver a guardar dependiendo del modo de salida. Los formatos de texto "pobres" como Texto (ANSI), etc. no almacenan gráficos.

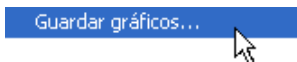


Aún así, con Readiris puede guardar gráficos sin llevar a cabo el reconocimiento de texto. Dado que el programa Readiris genera imágenes en blanco y negro, en escala de grises o en color, con él pueden capturarse dibujos y fotografías muy fácilmente.

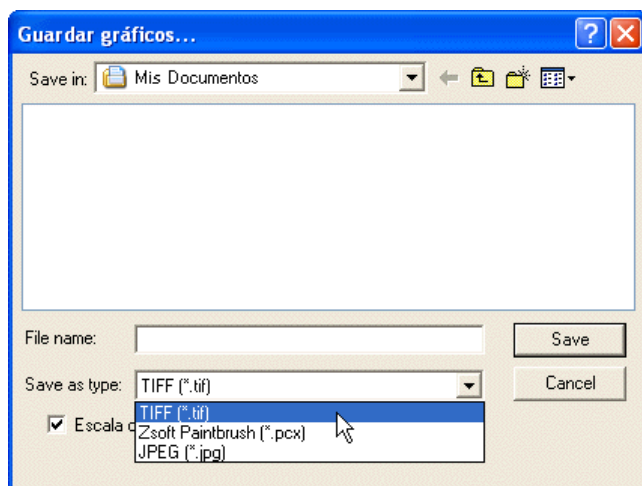
¿Cómo? Trace una zona de gráfico alrededor de las ilustraciones, caricaturas, etc. que desee capturar. Hágalo de la misma forma que traza una ventana de texto o de tabla, pero en este caso seleccione la herramienta "Ventana de Gráfico".



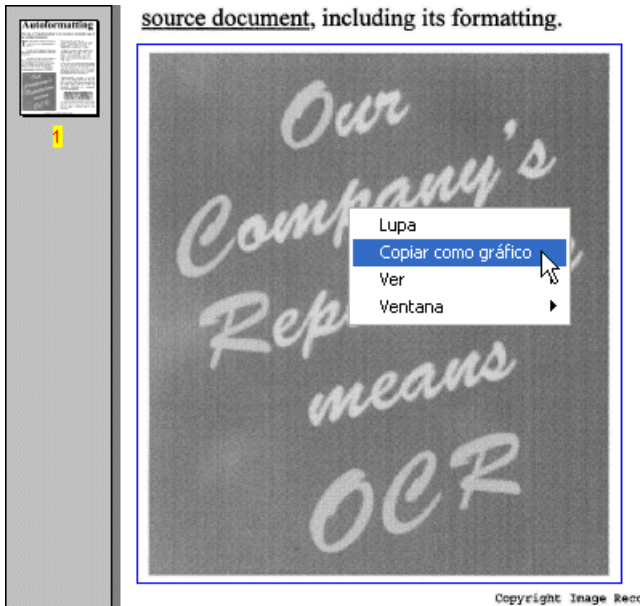
A continuación elija la opción "Guardar Gráficos" en el menú "Archivo".



Se le pedirá que especifique un nombre de archivo. Determine el formato gráfico que desee usar. Seleccione un formato que pueda utilizar con su *software* de retoque. Dispone de los formatos JPEG, TIFF y Paintbrush (PCX). Active la opción "Escala de Grises/Color" que permite guardar las imágenes bien en color, bien en escala de grises.



Para enviar un gráfico al portapapeles mejor que guardar un archivo de imagen, haga clic con el ratón sobre una ventana de gráfico y seleccione la opción "Copiar como Gráfico": la zona del gráfico que se haya seleccionado con el ratón está lista para ser pegada.



LECTURA DE FAXES Y RECONOCIMIENTO DIFERIDO

Al guardar imágenes en archivos graficos se abre otra posibilidad: puede guardar la página *completa* y realizar un **reconocimiento diferido** más adelante. Eso fue lo que hicimos con las imágenes escaneadas para nuestro curso.

Escanee simplemente el documento y seleccione la opción "Guardar Toda la Página como Imagen" del menú "Archivo". Guardará así todo el documento en una única página. Se le preguntará en qué formato desea guardarlo: TIFF o Paintbrush (PCX).

Guardar toda la página como imagen...
Guardar todas las páginas como imagen...

Seleccione la opción "Guardar Todas las Páginas como Imagen" para obtener un documento de múltiples páginas. En este caso, dispondrá de un formato: TIFF con múltiples páginas.

Ahora puede seleccionar el disco como origen de la imagen y abrir el archivo gráfico con el botón "Abrir" (o con la opción equivalente del menú "Acción"). (Si utiliza la opción "Abrir" del menú "Archivo", ni siquiera tiene que actualizar el origen de la imagen.)

Como el programa reconoce imágenes en color, en escala de grises y en blanco y negro, Readiris le permite abrir documentos PDF Adobe Acrobat y archivos de imágenes JPEG, Paintbrush (PCX), fax DCX (una versión con múltiples páginas del formato Paintbrush), PNG, TIFF (no comprimidos, comprimidos en LZW, packbits, Grupo 3 y Grupo 4), TIFF con múltiples páginas y mapas de bits de Windows (BMP).

Esta capacidad resulta especialmente útil para convertir sus **faxes** en archivos de texto editables. Readiris usa su máxima capacidad para reconocer faxes. El programa detecta las resoluciones típicas de fax - 100 x 200 dpi ("calidad normal"), 200 x 200 dpi ("calidad fina") y 200 x 400 dpi ("calidad super-fina") - y preprocesa automáticamente estas imágenes para asegurar unos resultados de OCR óptimos.

Sin embargo, de ser posible, pida a sus correspondientes que envíen sus faxes con calidad "fina", puesto que dichos faxes darán mejores resultados de OCR.

No olvide que puede abrir imágenes directamente desde el Explorador de Windows haciendo clic con el botón derecho del ratón en un archivo gráfico y seleccionando la opción "Reconocer" del menú "Contexto". También puede arrastrarlos con el ratón desde el Explorador de Windows hasta la zona de imágenes o sobre el icono de Readiris, y se abrirán automáticamente.

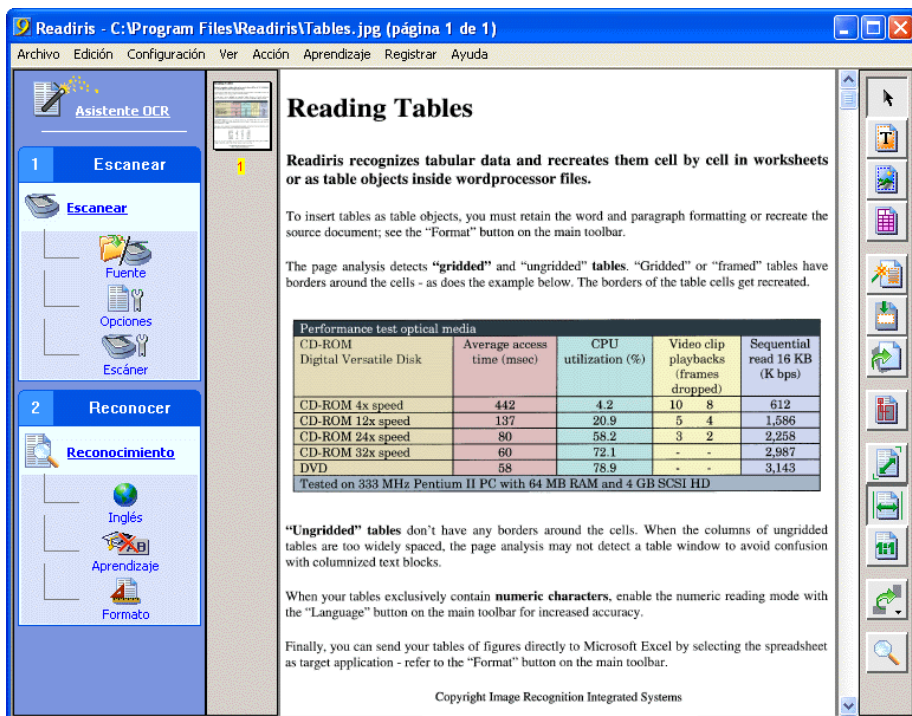


RECONOCIMIENTO DE TABLAS

Hasta ahora hemos reconocido textos y faxes y hemos guardado gráficos. Ahora procesaremos una tabla. Coja una tabla con cifras y escanéela o bien abra la imagen de muestra TABLES.JPG que se encuentra en la carpeta Readiris.

En realidad, la imagen TABLES.JPG contiene dos tablas ¡y no es una coincidencia! El análisis de página las divide en ventanas de tabla y Readiris las reconstruirá reproduciéndolas celda a celda en un programa de hoja de cálculo o insertando una tabla en el archivo del procesador de textos que usted prefiera.

Daremos algunos ejemplos. Empezaremos con la tabla con cuadrícula o “enmarcada”; es decir, la que tiene bordes alrededor de las celdas.



Para obtener una reproducción exacta de la tabla, ejecute el reconocimiento con la opción "Conservar el Formato de Palabras y Párrafos" o bien "Volver a Crear el Documento Original". Abra un procesador de textos y verá el resultado: las células y los bordes han sido creados por Readiris uno a uno! (Naturalmente, también puede incluir párrafos de texto normales en este archivo.)



table - Microsoft Word

File Edit View Insert Format Tools Table Window Help

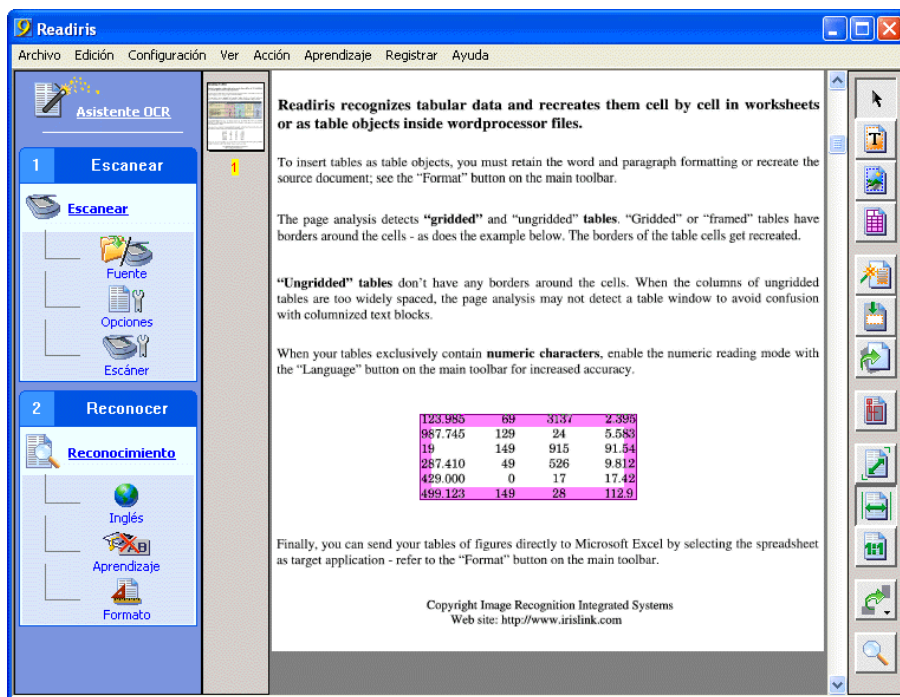
Type a question for help

Times New Roman 12 B I U

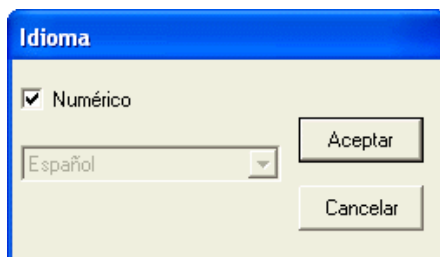
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

Performance test optical media					
CD-ROM	Average access	CPU	Video clip		Sequential
Digital Versatile Disk	time (msec)	utilization (%)	playbacks		read 16 KB
			(frames		(K bps)
			dropped)		
CD-ROM 4x speed	442	4.2	10	8	612
CD-ROM 12x speed	137	20.9	5	4	1,586
CD-ROM 24x speed	80	58.2	3	2	2,258
CD-ROM 32x speed	60	72.1	-	-	2,987
DVD	58	78.9	-	-	3,143
Tested on 333 MHz Pentium II PC with 64 MB RAM and 4 GB SCSI HD					

Pasemos ahora al ejemplo de las tablas sin cuadrícula; o sea, las que no tienen bordes alrededor de las celdas. ¡Observe que, pese a todo, el análisis de página detecta la tabla!

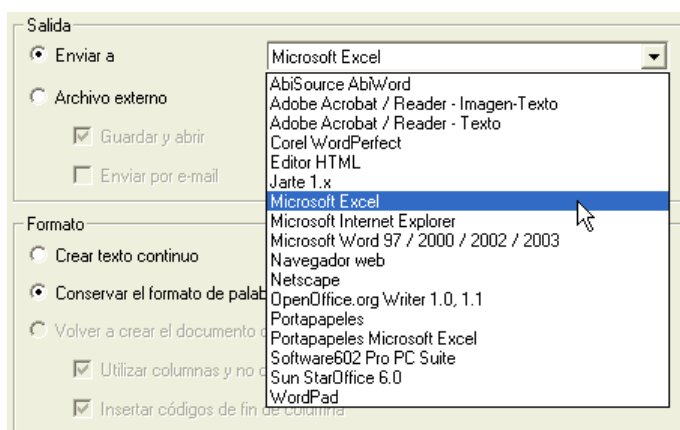


Para una precisión óptima de reconocimiento, limite el reconocimiento a los **símbolos numéricos** con el botón "Idioma". (El modo numérico no es estrictamente numérico, incluye los símbolos 0 a 9, +, *, /, %, , (coma), . (punto), (,), -, =, \$, £, ¥ y el símbolo €.)

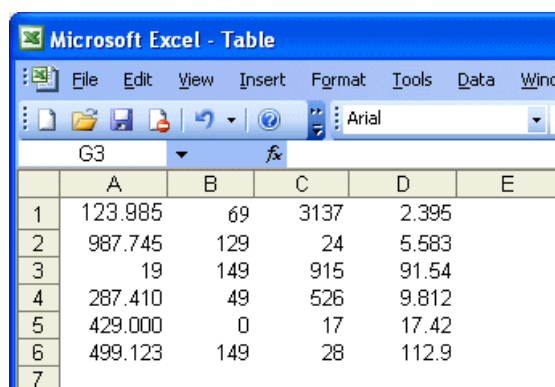


Puesto que esto sólo se puede hacer si la tabla no contiene símbolos alfabéticos - de lo contrario, las porciones de texto no se reconocerán correctamente - ahora podemos activar el modo numérico, algo que no podíamos hacer con la primera tabla.

En esta ocasión, enviaremos directamente el resultado del reconocimiento al programa de hoja de cálculo Microsoft Excel. Seleccionaremos, pues, Excel como aplicación de salida mediante el botón "Formato".



La hoja de cálculo se inicia automáticamente y el resultado es idéntico al de la típica estructura de tabla con filas y columnas, en la que podrá procesar datos inmediatamente.



	A	B	C	D	E
1	123.985	69	3137	2.395	
2	987.745	129	24	5.583	
3	19	149	915	91.54	
4	287.410	49	526	9.812	
5	429.000	0	17	17.42	
6	499.123	149	28	112.9	
7					

Puede ocurrir que el análisis de página a veces no detecte como tal alguna tabla sin cuadrícula y por tanto no la sitúe en una zona de tabla. Esto se debe a que el espacio entre columnas es muy amplio. Readiris pretende de esta manera evitar la posible confusión con bloques de texto en los que haya columnas y no tablas. Para trazar una ventana de tablas manualmente, haga clic en la herramienta "Ventana de Tabla" de la barra de herramientas de imágenes y siga el procedimiento habitual. Una vez más, la etiqueta emergente del botón le indicará el número de ventanas de tabla.

Definir una ventana de tabla: 0

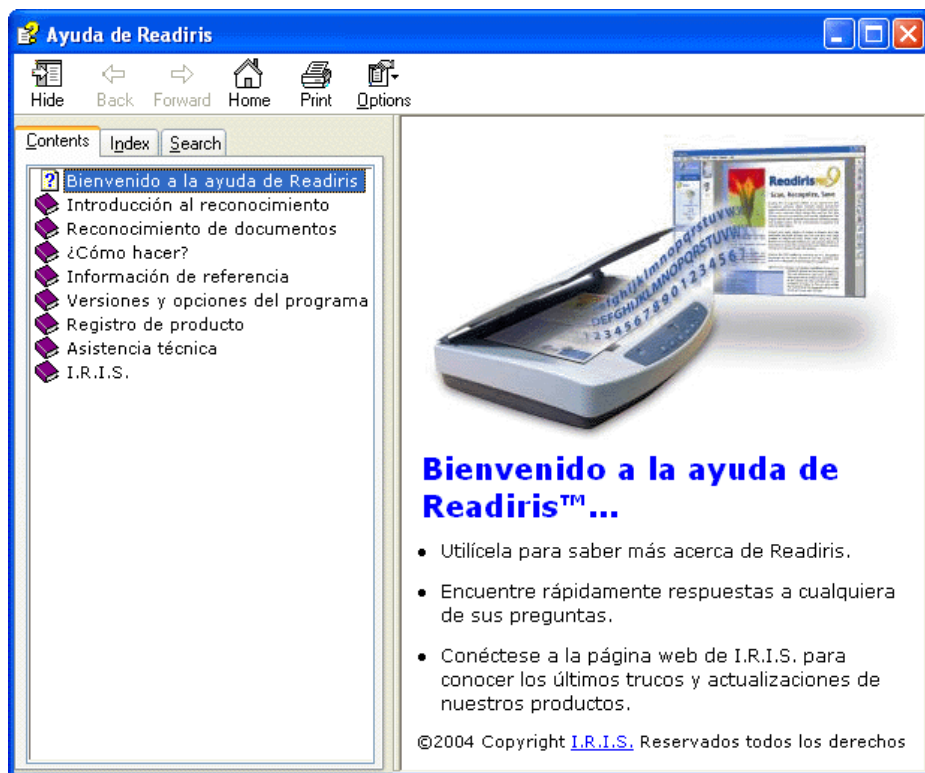




CÓMO OBTENER AYUDA EN LÍNEA

Con esto concluimos nuestra sinopsis general de Readiris. Es posible que este manual no contenga información de último minuto. Por ello le recomendamos que consulte el sistema de ayuda en línea para más información sobre Readiris.

Para ello vaya al menú "Ayuda". La opción "Temas de Ayuda" y su macro, la tecla F1, le permiten navegar a través de los numerosos temas de ayuda.



El resto de opciones del menú "Ayuda" le indican cómo obtener asistencia técnica para el producto, cómo ponerse en contacto con I.R.I.S., cómo acceder directamente a la página web de I.R.I.S., etc.

