

Capítulo 2

VISITA GUIADA

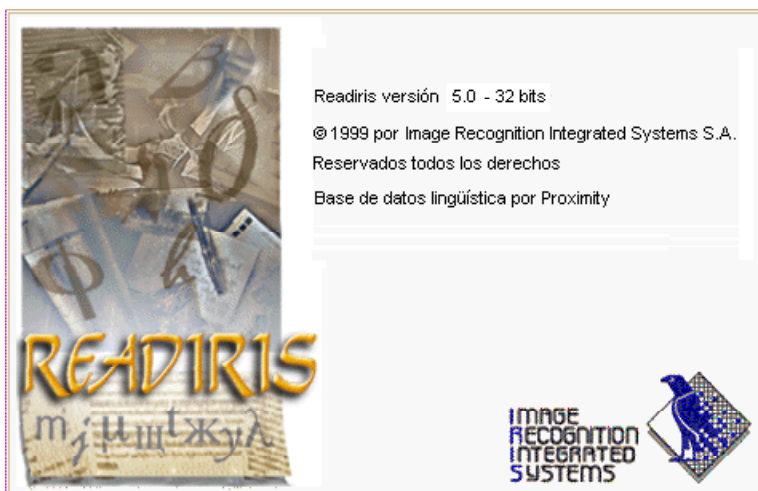
Readiris es un paquete con la tecnología OCR más moderna provisto de numerosas características avanzadas. En este capítulo describiremos todas sus principales características y añadiremos consejos sobre el uso de Readiris.

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.



Aparecerán la pantalla de inicio y la ventana de la aplicación Readiris. La pantalla de inicio muestra la versión y los derechos de autor del *software* de Readiris. Esta pantalla desaparecerá haciendo clic en cualquier parte de la pantalla.



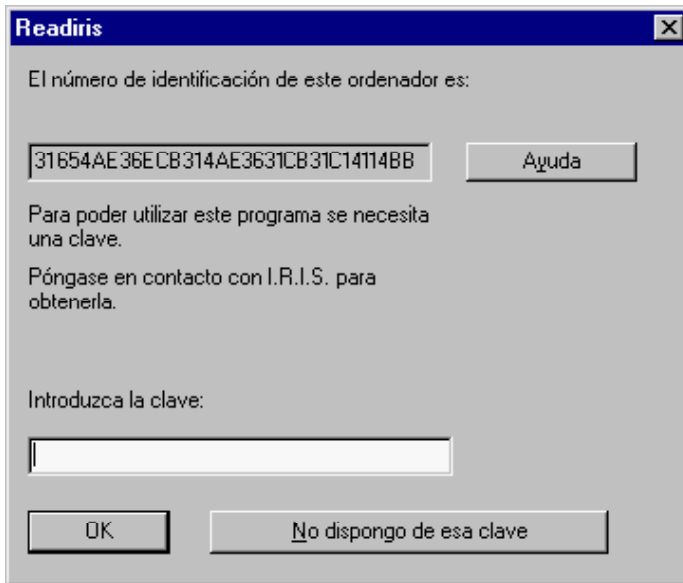
La siguiente ventana se refiere al asistente de OCR. Haga clic en Cancelar.

PRIMERA EJECUCIÓN

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

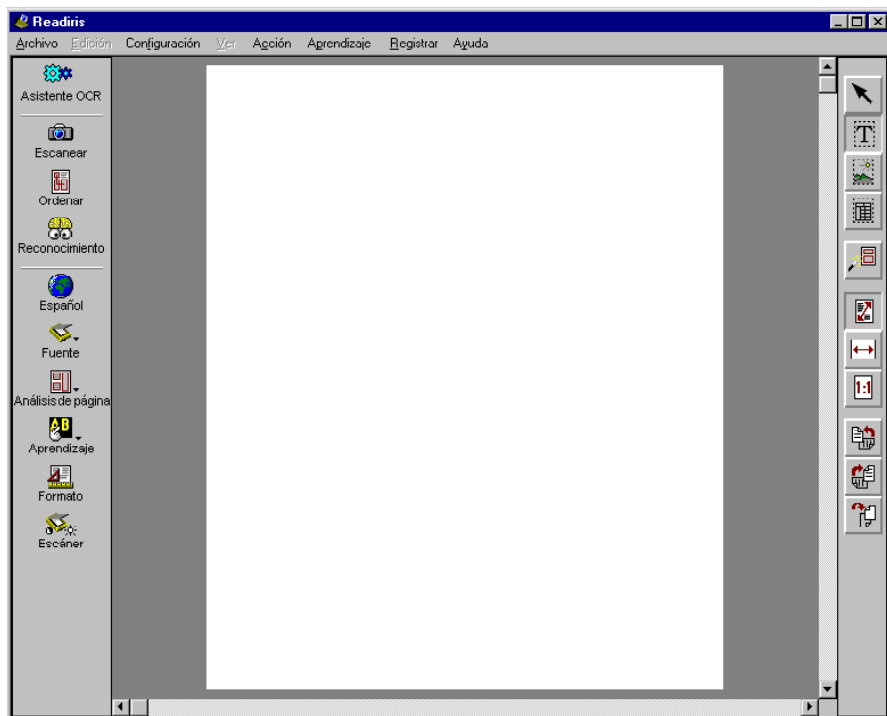
En tal caso, la utilización de Readiris está limitada a un mes y, al registrarse, recibirá de I.R.I.S. una **clave de acceso** gratuita para continuar utilizando el *software* transcurrido dicho plazo.

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



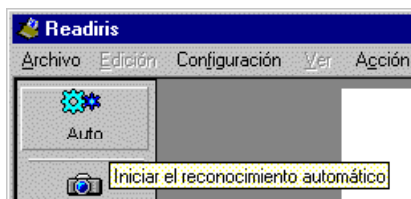
DESCUBRIR EL INTERFAZ DE READIRIS

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 gráficas** 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 y un **mensaje de herramienta** le dirá lo que hace el botón.



Por último, la parte interior de la ventana o **zona de imágenes** es donde se muestran las imágenes escaneadas. Puede arrastrar los archivos gráficos hasta la zona de imágenes para reconocerlos.

PRIMERA LECCIÓN INTRODUCTORIA

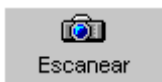
La mejor manera de familiarizarse con el manejo de Readiris es sin duda utilizándolo. El *software* contiene una serie de **imágenes ya escaneadas** que le permitirán comenzar incluso si no tiene un escáner conectado a su ordenador. Pasemos a estas imágenes.

El botón “Fuente” de la barra de botones determina si va a utilizar un escáner o una imagen ya escaneada como origen de la imagen. Readiris le permite abrir archivos TIFF (sin compresión, paquetes de bits, Grupo 3 y Grupo 4 comprimidos), Paintbrush (PCX) y mapas de bits de Windows (BMP). Sólo pueden utilizarse imágenes en blanco y negro. Esta función resulta especialmente útil para convertir sus **faxes** en archivos digitales de texto.

Puesto que vamos 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.



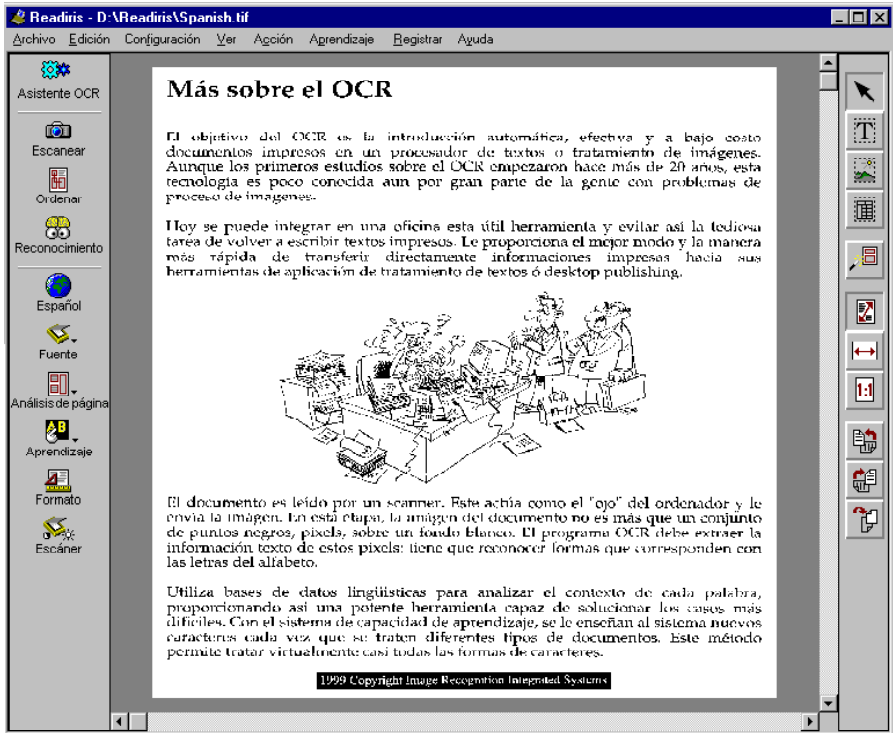
A continuación haga clic en el botón “Escanear”.



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, a la opción “Escanear” del menú “Acción” la sustituye la opción “Abrir”).)



Se le pedirá que elija un archivo gráfico. Seleccione el archivo SPANISH.TIF de la carpeta Readiris. La imagen es leída en el disco y aparece progresivamente en la zona de imágenes.



Otra forma de abrir imágenes ya escaneadas es el **“arrastre con el ratón”**. Arrastre un archivo desde el Explorador de Windows hacia la zona de imágenes de Readiris y éste se abrirá inmediatamente.

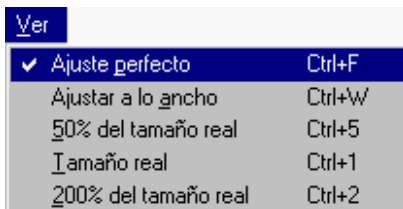
La barra de herramientas gráficas situada a la derecha de la ventana de aplicación de Readiris contiene todas las opciones que necesita durante la vista preliminar de la imagen: herramientas para indicar las zonas de interés, para girar la imagen, para ampliarla o reducirla, etc.



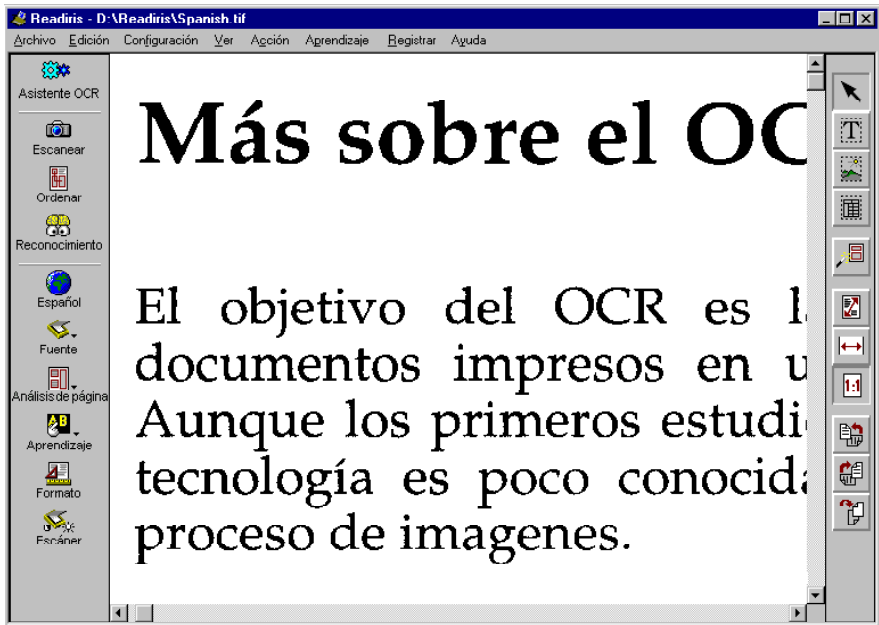
AMPLIACIÓN DE IMÁGENES

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

La barra de herramientas gráficas contiene botones que le permiten ampliar a tamaño real, ajustar la imagen al ancho de página y de 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.)



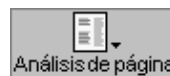
Por último puede pulsar el botón *derecho* del ratón en una zona de la imagen escaneada para ampliarla inmediatamente a tamaño real. Vuelva a pulsar el botón derecho del ratón para volver al tamaño anterior.



UNO, DESCOMPONER UNA IMAGEN ESCANEADA

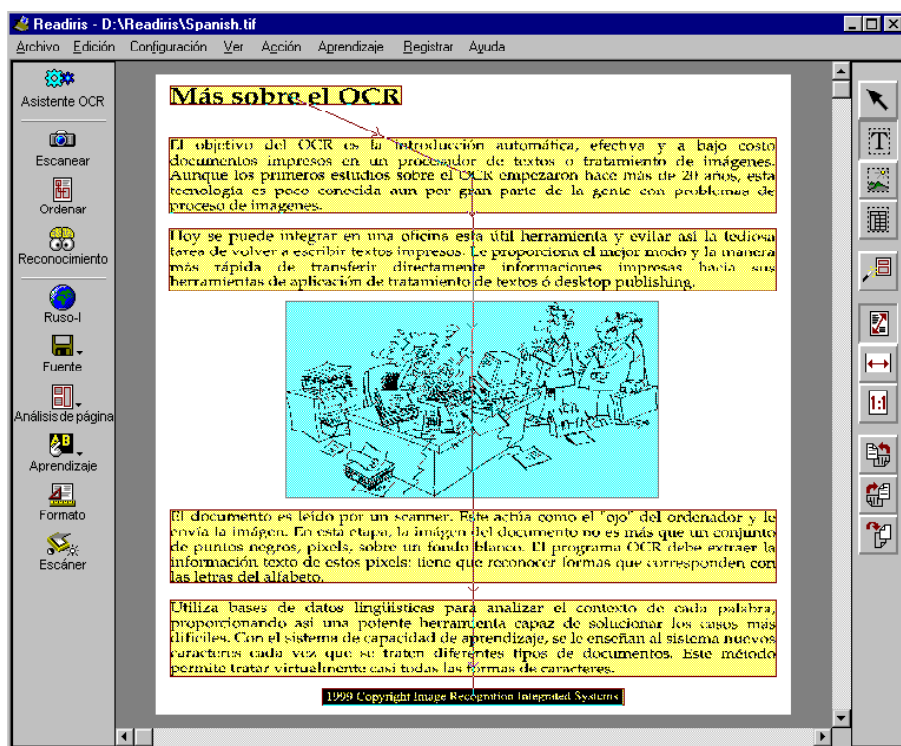
Una vez escaneada la imagen debe indicar las partes que desea convertir en texto modificable trazando marcos, denominados “ventanas”, alrededor de las zonas de interés.

De hecho, Readiris lo hará automáticamente si activa la opción “Análisis de Página” de la barra principal de herramientas.





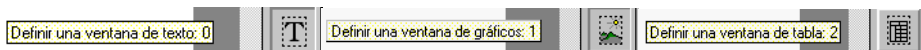
La descomposición automática de páginas resulta especialmente útil para reconocer **textos con columnas** y documentos con un diseño de página complejo (incluirán posiblemente gráficos y tablas).



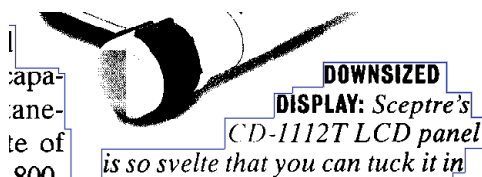
La división de página se realiza mediante tres **tipos de ventanas**: de texto, de gráficos y de tablas. Readiris separa en cada página bloques de texto, zonas de tablas y zonas de gráficos (con fotografías, ilustraciones, etc.). (La manera de reconocer tablas y guardar gráficos se tratará ampliamente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

Un **color** indica el tipo de ventana de que se trata: las de texto son amarillas, las de gráficos son azules y las de tablas, moradas.

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. (De esta forma, y gracias al escáner, podrá generar imágenes totalmente invertidas de páginas que contengan texto blanco sobre fondo negro. Ver más adelante.)

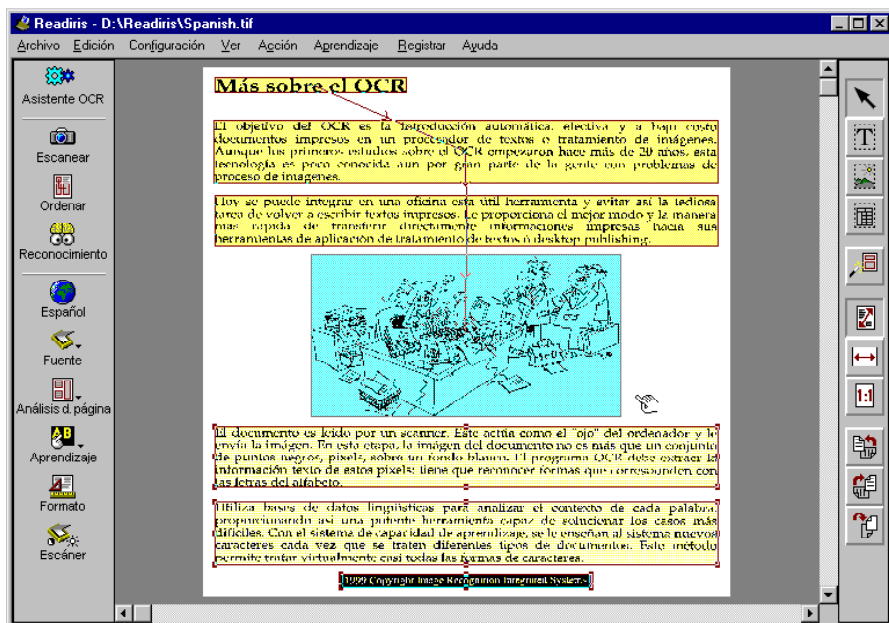
UNA Y MEDIA, CLASIFICACIÓN DE VENTANAS

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 principal. 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 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

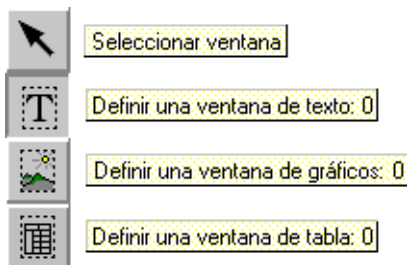
análisis de página – simplemente haga clic en el botón “Analizar Página” de la barra de herramientas gráficas.



(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 ha de leerse de arriba a abajo y de derecha a izquierda.)

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.



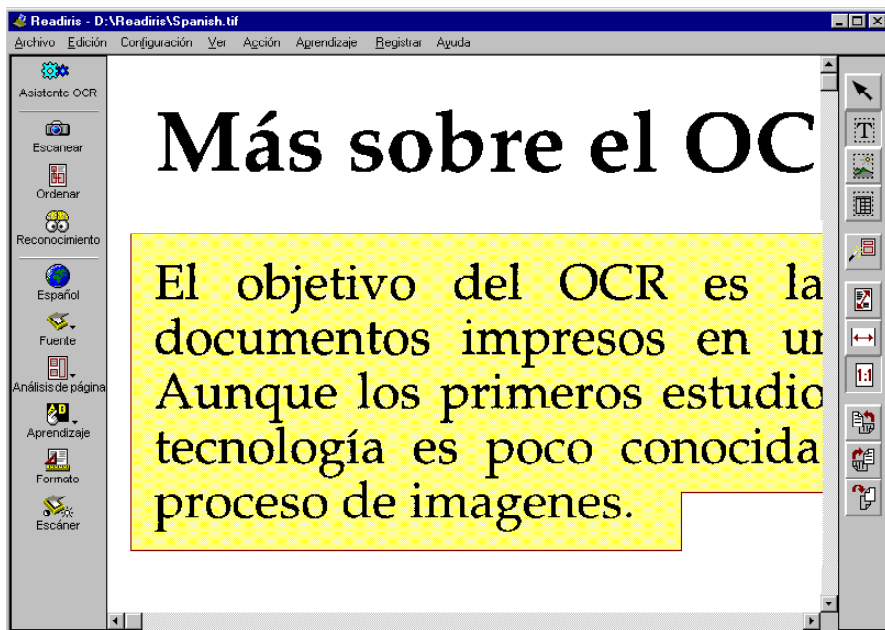
Para **crear** un rectángulo alrededor de zona concreta, seleccione la herramienta correspondiente en la barra de herramientas gráficas, coloque el cursor en la parte superior izquierda de la ventana que desee trazar, pulse el botón izquierdo del ratón, extienda la ventana moviendo el ratón hacia la esquina inferior derecha



de la ventana y vuelva a pulsar el botón izquierdo. (No se permiten lados inferiores a 1 mm., pues no podrían contener ni un solo carácter.)

Las ventanas se clasifican automáticamente en el orden de su creación: unas flechas 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!

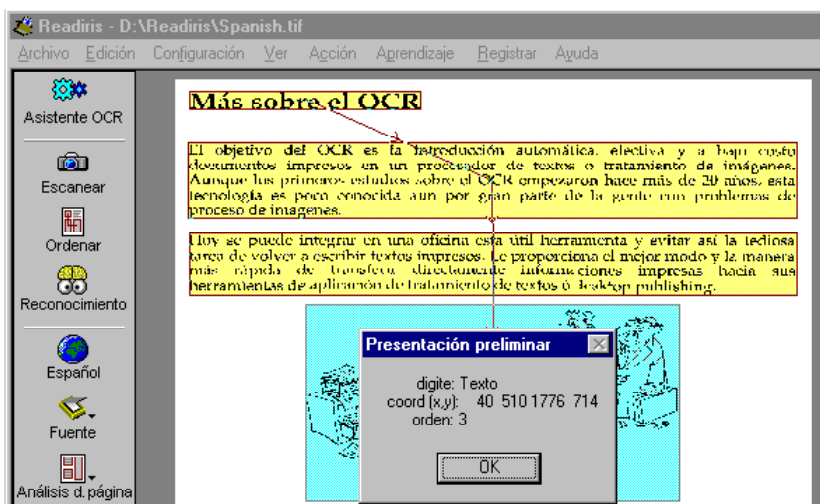
Para modificar, mover y eliminar ventanas, primero debe **seleccionarlas**. Para ello, seleccione la herramienta “Selección de Ventana” (la “flecha”) en la barra de herramientas gráficas y haga clic en el interior de la ventana. Aparecerán unas marcas rectangulares en cada esquina y en el centro de los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



Más sobre el OCR

Para **deseleccionar** ventanas, pulse el botón izquierdo 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 Control y haga clic en la ventana principal.

Nota: puede mostrar el tipo, coordenadas de píxeles y orden de clasificación de una ventana pulsando la tecla Alt mientras la selecciona.



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 las opciones “Cortar” o “Borrar” del menú “Edición”. La opción “Cortar” corta la(s) ventana(s) y la(s) envía a una memoria temporal interna; la opción “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, *todas* las opciones 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.



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 Control+A y hacer clic en la opción “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 Alt+B.

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 tareas.



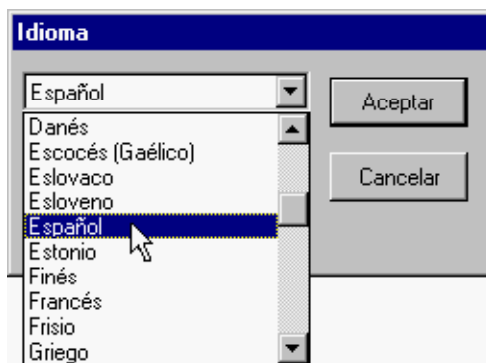
READIRIS LO LLEVA POR TODO EL MUNDO

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 las propiedades de idioma y 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.



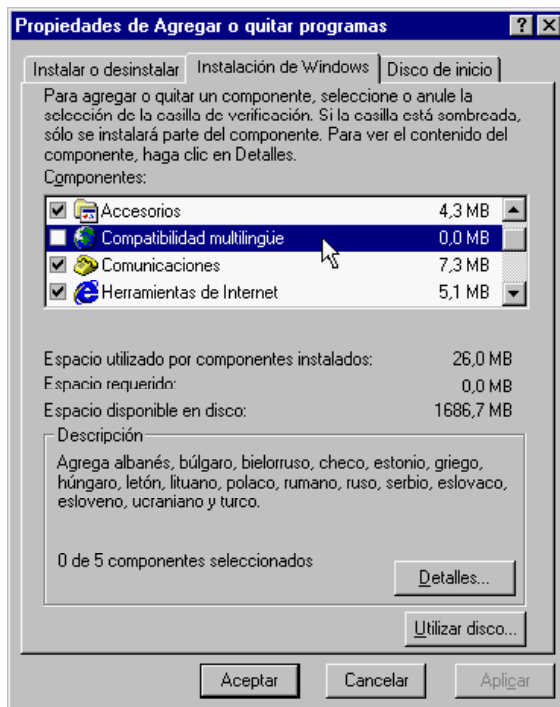
Para pasar directamente a un idioma pulse la letra inicial de éste: si está seleccionado el español y desea localizar el húngaro, pulse la letra “H” para ir directamente hasta él. Si varios idiomas tienen la misma inicial, pulse la letra varias veces para recorrer las opciones. Por ejemplo, Readiris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

Readiris no se limita ni con mucho al inglés: se pueden utilizar hasta 55 **idiomas**. Se pueden utilizar todos los idiomas americanos y europeos, incluyendo los caracteres centroeuropeos, griegos, turcos, cirílicos (“rusos”) y bálticos.

Opcionalmente, puede reconocer documentos escritos en **lenguas asiáticas**. El *software* opcional “Módulo OCR asiático” permite el reconocimiento del japonés y del chino simplificado, no así del chino tradicional. (El chino simplificado se habla en el continente y el chino tradicional lo utilizan las “5 Grandes” comunidades - Hong Kong, Taiwan, Singapur, etc.)

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

Nota: se requiere la configuración adecuada de Windows para mostrar en pantalla caracteres centroeuropeos, griegos, turcos, cirílicos y bálticos. Quizá deba instalar el “paquete de servicio” de Windows “Compatibilidad Multilingüe” para que su sistema pueda procesar dichos idiomas. Seleccione el icono “Agregar o Quitar Programas” de “Panel de Control” para ver si dicho módulo de Windows está instalado. Dependiendo del paquete de *software* que haya adquirido, el CD-ROM de Readiris puede contener información detallada sobre la forma de instalar este módulo de Windows.



Para visualizar y editar documentos en lenguas asiáticas, deberá instalar una versión asiática (japonesa o china) de Windows. También podrá utilizar programas de “emulación” especializados (como por ejemplo UnionWay AsianSuite o TwinBridge AsianBridge) en una versión occidental de Windows, para representar correctamente los ideogramas de las lenguas asiáticas.

La selección del idioma adecuado para el documento es imperativa. En función de la selección del idioma, el *software* sabe qué símbolos debe reconocer. El

soporte multilingüe garantiza un reconocimiento correcto de caracteres “exóticos”, como ç, ß, γ y ø.

Asimismo, el *software* 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, por no hablar de 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 “l”. 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

La imagen muestra una línea de texto con varias formas de la letra 'l' y el número '1'. La primera línea contiene '193', '1950s.', y '1hr'. La segunda línea contiene 'Well,' y 'Rossellini'. Las formas de 'l' y '1' en la segunda línea son ambiguas y se parecen a las de la primera línea.

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 los alfabetos** griego, cirílico, cualquiera de los asiáticos y el latino.

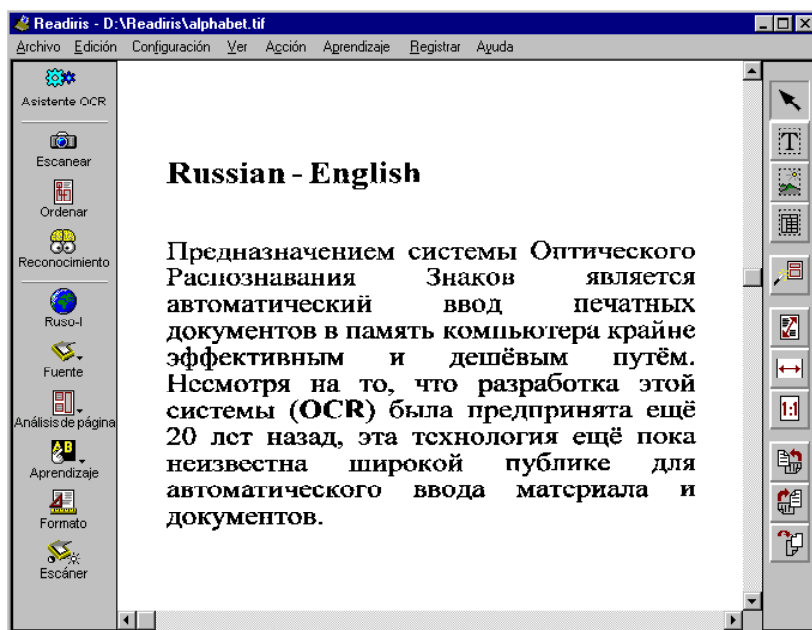
Asegúrese de seleccionar “Griego-Inglés” o la propiedad adecuada de cirílico – por ejemplo “Bielorruso-Inglés”. En otras palabras, si selecciona únicamente



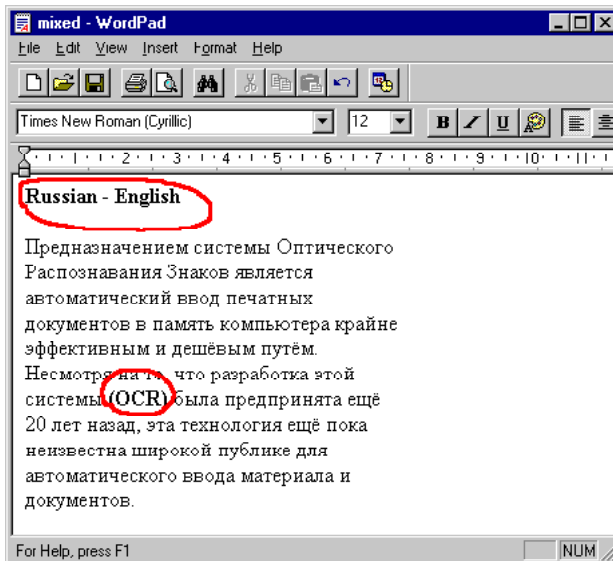
“Griego” o “Bielorruso” como idioma del documento, no espere que salgan bien los caracteres latinos.



He aquí un ejemplo en el que un texto contiene algunas palabras inglesas – abra el archivo gráfico ALPHABET.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 **fuente** 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 tamaño. (Estas opciones no están disponibles para los documentos de lenguas asiáticas.) Expliquemos lo que esto significa.



Primero que nada, indique si va a reconocer documentos “normales” o impresos con **matriz de agujas** con la opción “Tipo de Fuente” en el menú “Configuración”.

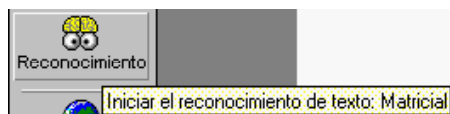


Es obligatorio establecer correctamente este parámetro: los símbolos “borrador”, también denominados “9 agujas”, están formados por puntos aislados y separados y su reconocimiento requiere rutinas altamente especializadas.

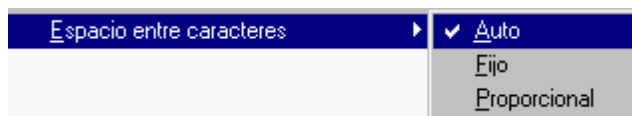
ape-descended life

La impresión con matriz de agujas “calidad de carta”, también denominado “25 agujas”, requiere propiedades normales, al igual que las **calidades de impresión** imprenta, mecanografiado, impresión láser e impresión con chorro de tinta. Obviamente, “Normal” 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 de “monoespacio”), 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. Como 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.

Courier, fuente fija
Letter Gothic, fuente fija

Century Schoolbook, fuente proporcional
Arial, fuente proporcional

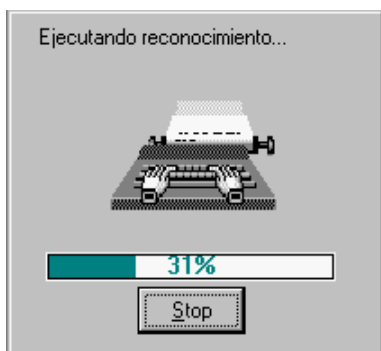
La solución más sencilla es dejar esta opción siempre en el valor por defecto “Auto”, que significa que Readiris detectará automáticamente el tamaño de la fuente.

READIRIS SE HACE CADA VEZ MÁS INTELIGENTE

Una vez seleccionado el idioma del documento y establecidas las características del documento, puede pulsar el botón “Reconocimiento”.



El avance del reconocimiento se indica en la pantalla. Puede pulsar el botón “Stop” (parar) para cancelar 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 “Aprender” de la barra principal de herramientas. El aprendizaje interactivo está activado por defecto.

(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 em-

bargo, 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 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 de caracteres. (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.



Aprendizaje

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 dos caracteres unidos (“ligadura”) y pulsamos “Aprendizaje” 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 “Aprendizaje” 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 “c”, por lo que debe pulsar el botón “No Aprender” a fin de no confundirlo con el símbolo “e”.

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 nueve 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.

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

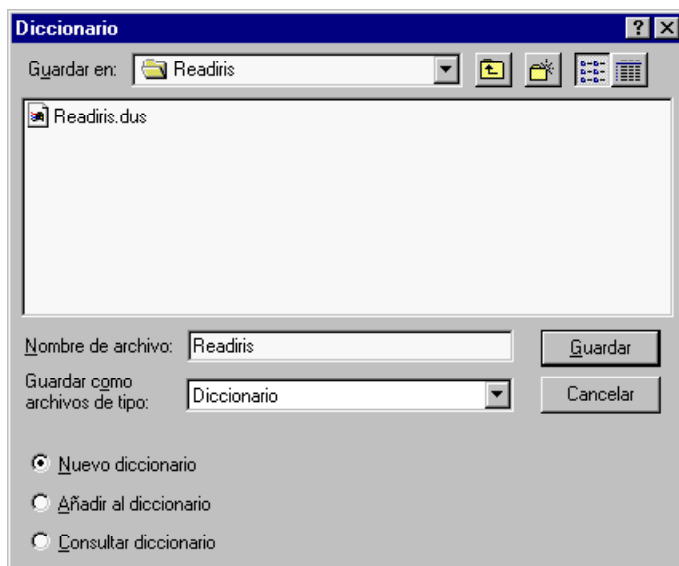
Los resultados de cada sesión de aprendizaje se conservan temporalmente en la memoria del ordenador, pero pueden y deben guardarse en archivos denominados “diccionarios” para su uso ulterior.

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



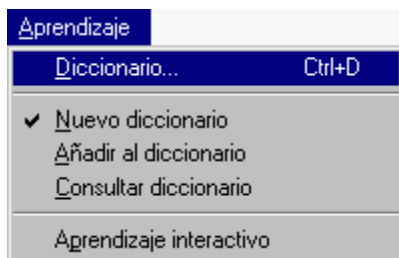
estas bibliotecas de fuentes. Podría decirse que Readiris se vuelve más inteligente cada vez que la usa.

¿Cómo funciona esto? El funcionamiento de los diccionarios de fuentes está controlado por el menú “Aprendizaje”: tiene que seleccionar un diccionario con la opción “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. Para mayor claridad le recomendamos que utilice nombres significativos para los diccionarios de fuentes, por ejemplo INFORME.DUS, PALATINO.DUS etc. 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.)

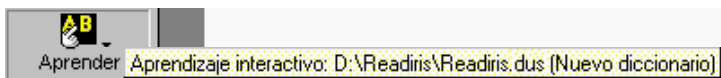
El modo añadir indica que los resultados del aprendizaje se guardarán en diccionario *existente*: el reconocimiento utiliza la información que ya contiene el diccionario y se le añaden nuevas formas de fuentes. En otras palabras, esta opción le permite elaborar un diccionario de fuentes en varios pasos.

(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.)

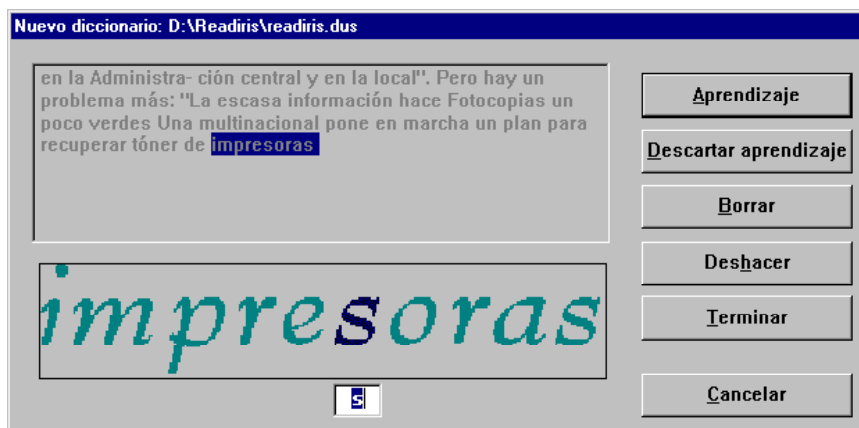
Con la última opción “Consultar Diccionario”, el diccionario funciona en modo de sólo lectura y se aprovecha el diccionario *sin* añadirle nuevas formas.

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. Si desea reconocer documentos de varias páginas, al escanear la siguiente página, el diccionario de usuario entra automáticamente en modo “añadir”.

El mensaje de la herramienta “Aprender” 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 DIRECTAMENTE A LA APLICACIÓN DE TRABAJO

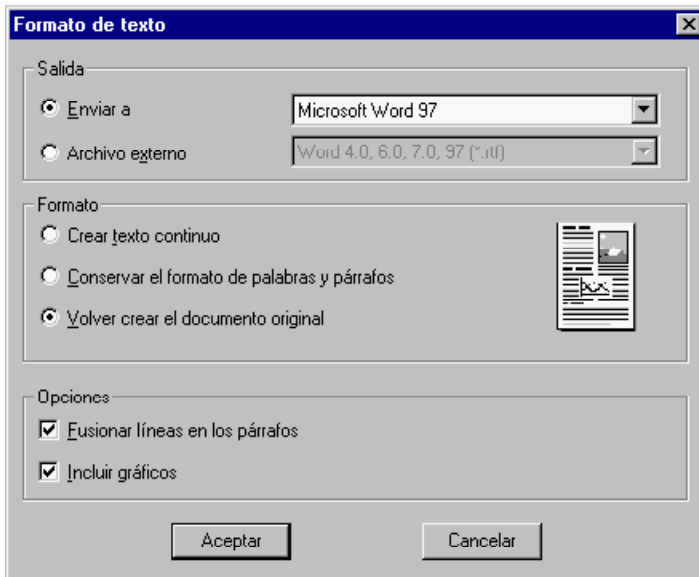
Con el aprendizaje interactivo concluye el reconocimiento de caracteres. Si Microsoft Word 97 es el programa receptor de datos por defecto, el procesador de textos se ejecuta automáticamente (si fuese necesario) una vez concluido el reconocimiento, y se inserta el texto reconocido.

La imagen escaneada vuelve a aparecer con la división en zonas inicial para su procesamiento ulterior, y permanecerá en pantalla hasta que escanee otra página.

De hecho, ha convertido un documento de papel en un archivo digital modificable, en un tiempo entre 10 y 25 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.

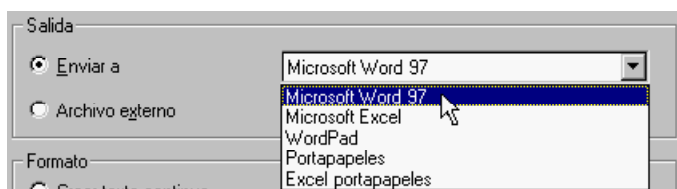
En realidad, Readiris le ofrece tres métodos diferentes a la hora de guardar el resultado del reconocimiento: por un lado, enviar directamente el documento reconocido a una aplicación; por otro, guardar el resultado en un archivo externo; y, por último, copiar el resultado en el portapapeles de Windows.

El **formato 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 reconocimiento entre el escáner y las aplicaciones para Windows. ¡Con ella podrá **enviar** directamente el documento que acaba de escanear a un procesador de textos o a una hoja de cálculo! Readiris exporta directamente los documentos reconocidos a Microsoft Word 97, Microsoft Excel o WordPad.



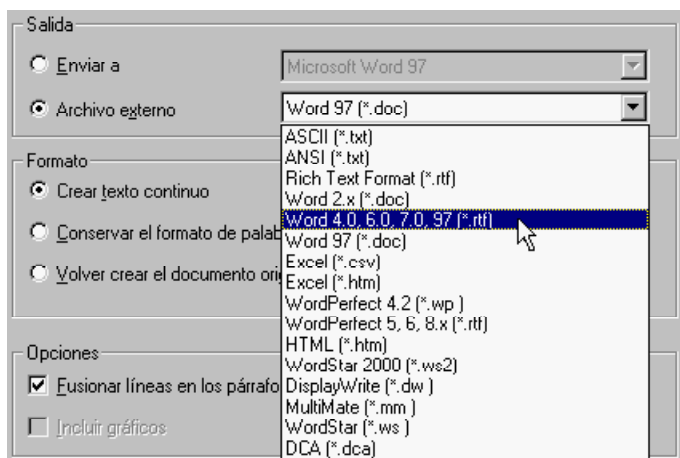
Una vez concluido el reconocimiento, se ejecuta la aplicación de destino y el documento reconocido se visualiza en un nuevo archivo de texto o en un libro de trabajo.

Por favor, espere mientras se carga
Microsoft Word 97 ...

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.

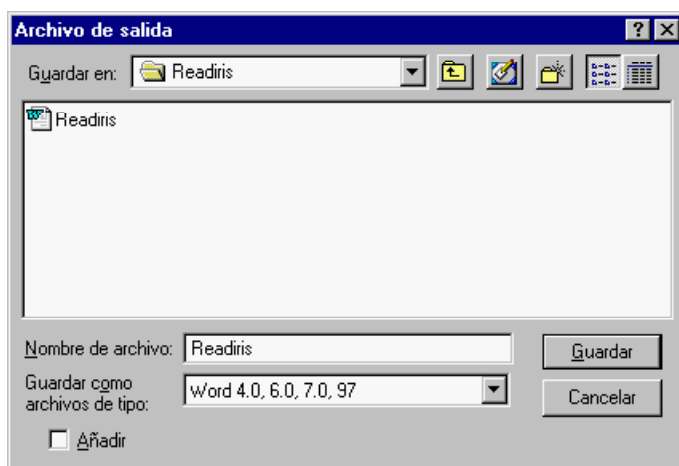
GUARDAR LOS RESULTADOS EN UN ARCHIVO DE TEXTO

Puede, también, guardar el resultado del reconocimiento en un archivo “externo”. Readiris reconoce la gran variedad de formatos que ofrecen los procesadores de texto y hojas de cálculo más conocidos - Microsoft Word (DOC), RTF, HTML, etc.

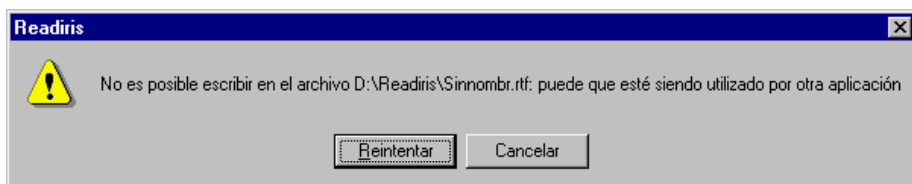


Al seleccionar la opción “Archivo externo”, se le pedirá que guarde el texto reconocido una vez concluida la fase de reconocimiento.

Puede crear un nuevo archivo y añadirle otro archivo de texto ya existente - para ello, dispone de la opción “Añadir”. Puede, por ejemplo, añadir dos páginas de texto escaneadas a cualquier trabajo que haya redactado usted mismo. ¡Si añade un archivo ya existente, asegúrese de que se encuentra abierto, para evitar así sobrescribirlo!)

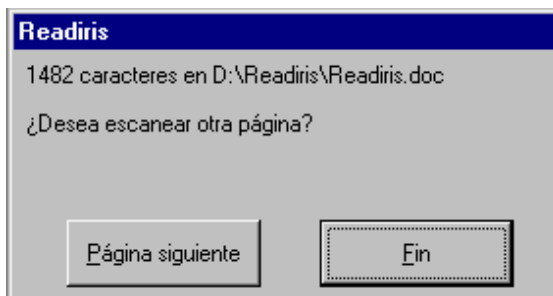


Puede crear un nuevo archivo o añadir el texto a un archivo de texto existente - para ello tiene el botón “Añadir”. Por ejemplo, puede añadir dos páginas escaneadas de texto a un estudio de su propia creación. Si desea agregar el texto a un archivo existente, asegúrese que no está abierto, ya que en tal caso no podrá hacerlo.



RECONOCIMIENTO DE VARIAS PÁGINAS

Si hace clic en “Guardar” para guardar el texto, el sistema le preguntará si desea reconocer otra página. (Si envía el texto reconocido a una aplicación o al portapapeles, obviamente no se le pedirá que guarde el texto.)



En esta fase se presentan algunas **estadísticas** breves sobre el número de caracteres leídos.

Ahora puede abrir el texto reconocido con su tratamiento o editor de textos, importarlo a su programa de publicación o a cualquier otra aplicación basada en textos.

Nuestra lección para páginas sueltas acaba aquí, ¿pero cómo se guarda el texto de las páginas adicionales? O en otras palabras, ¿cómo se procesan los documentos formados por varias páginas? Realmente es muy sencillo, haga clic en “Página Siguiente” cuando se le pregunte si desea reconocer otra página. Al hacerlo se escaneará una nueva página si el escáner es el origen de la imagen. Segundo, el archivo resultante y el diccionario de fuentes entran automáticamente en modo “añadir”, de forma que puede agregar el texto y continuar cómodamente con el aprendizaje de fuentes.

Puede continuar el reconocimiento de otras páginas: el nuevo texto se añade al archivo de texto cada vez que lo guarde tras su reconocimiento.

ORGANIZACIÓN DEL TEXTO RESULTANTE

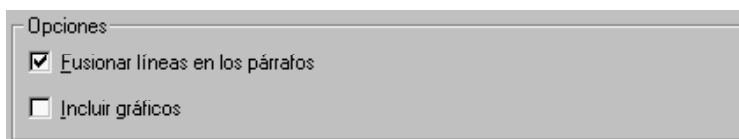
Guardar o exportar el texto requiere algo más que indicar simplemente un método de exportación o un nombre para el archivo resultante. También debe seleccionarse un formato de archivo y determinarse la apariencia del texto



reconocido. Es decir, hay que decidir adónde desea transferir 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” que se encuentra en el botón “Formato”.



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 (códigos 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.

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

CONFIGURACIÓN DEL ESCÁNER

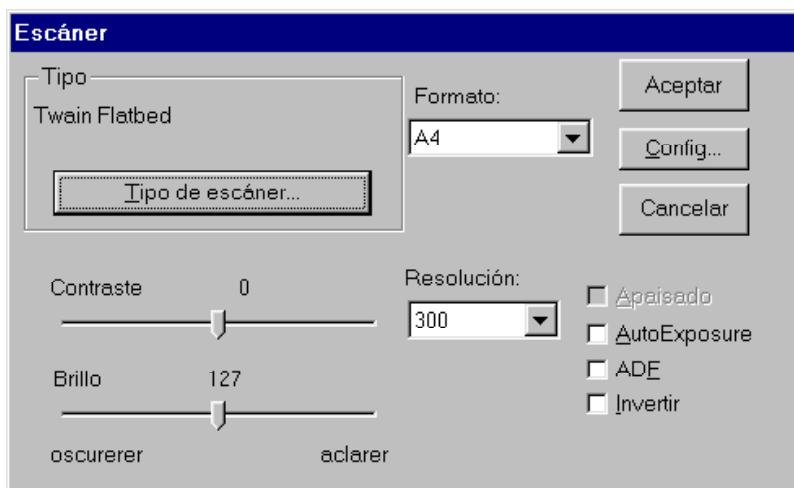
Ahora debe configurar su escáner. Se supone que el escáner y los controladores necesarios han sido instalados correctamente.

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**.

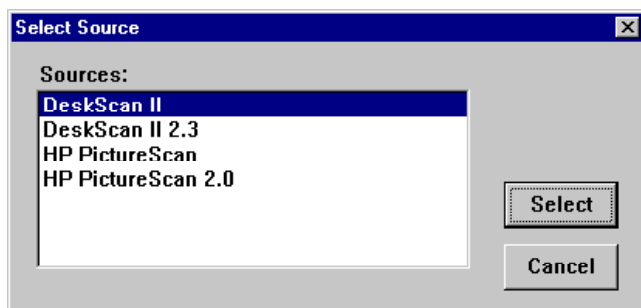


Si selecciona “<Image>” (Imagen) como “escáner” – las imágenes ya escaneadas funcionan como origen de las imágenes en todo momento – no tiene



que seleccionar el disco como origen de imagen con el botón “Fuente” de la barra principal de herramientas.

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



Su escáner también viene con una **plataforma de escaneado**, como Visioneer PaperPort o HP Document Assistant. Dependiendo del paquete de *software* que haya adquirido, en el CD-ROM de Readiris también puede encontrar manuales especializados sobre las plataformas de escaneado.

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.

Seleccione una **resolución** de 300 ppp para las aplicaciones normales, utilice una resolución más elevada (400 ppp) para la letra pequeña (menos de 10 puntos) y si el documento está degradado.

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

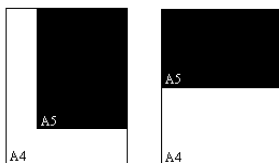
6 puntos

72 puntos

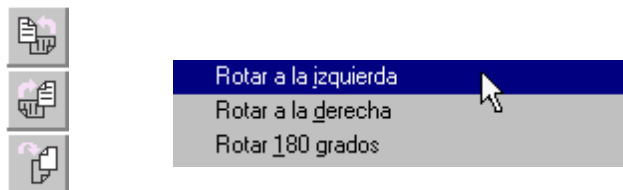
Readiris reconoce asimismo “**letras descendentes**” (“drop letters”), mayúsculas de gran tamaño que cubren varias líneas. (Desde luego, éstas no pueden ser superiores a 72 puntos.)

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

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.



Cada vez que necesite ajustar la orientación de la página, puede usar las herramientas de **rotación** de la barra de herramientas gráficas. (Las opciones correspondientes se encuentran en el menú “Ver”.)



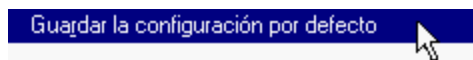
Existen tres direcciones de rotación: a la izquierda, la derecha e inversión. La rotación tarda algunos segundos en actualizar la imagen en sí y no sólo la que se muestra en pantalla.

Establezca el **brillo** y, si dispone de él, el **contraste**. El ajuste de contraste sólo existe en algunos escaners. Arrastre la barra deslizante para oscurecer o aclarar el documento; los valores seleccionados se representan en forma de números.

Por último, la opción “Invertir” le permitirá generar **imágenes “invertidas”**. Active esta opción si se dispone a procesar páginas enteras en las que aparezca texto blanco sobre fondo negro.

CÓMO GUARDAR LAS PROPIEDADES PREDETERMINADAS

Establezca correctamente todos los parámetros de escaneo 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.

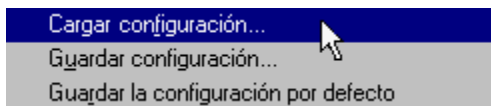


Los archivos de configuración no sólo contienen la **configuración** del escáner, también determinan si desea utilizar el aprendizaje interactivo, qué diccionario desea utilizar (y su modo de uso), el idioma y el tipo de letra de los documentos – por ejemplo, una letra proporcional normal–, el modo de salida y las opciones de formato que están en uso –por ejemplo, texto normal enviado 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.



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



CÓMO ACTIVAR EL ASISTENTE

Con el escáner ya configurado podemos pasar 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 de OCR” de la barra de herramientas principal (o seleccione la opción “Asistente de 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 “Ejecutar el asistente al iniciar”, que aparece en la primera pantalla del asistente. También puede desactivarlo usando la misma opción del menú “Configuración”.

RECONOCIMIENTO AUTOMÁTICO DE PÁGINAS

El asistente de OCR es una manera semiautomática de reconocimiento de páginas; pero Readiris también dispone de un sistema *totalmente* automático de reconocimiento. En vez de ir ajustando uno a uno todos los parámetros, lo que haremos será reconocer la página inmediatamente, sin ningún tipo de interrupción.

Para volver a situar el botón “Asistente de OCR” en la barra de herramientas principal junto al botón “Auto”, active la opción “Situación Asistente en Barra de Herramientas”, que se encuentra en el menú “Configuración”.



✓ Situar asistente en barra de herramientas



Seleccione el escáner como el origen de la imagen. Emplee para ello el botón “Fuente”. Posteriormente, haga clic en el botón “Auto” (o seleccione la opción “OCR Automático” del menú “Acción”).



El **reconocimiento automático** de caracteres significa que una página es escaneada, dividida en ventanas mediante análisis de página o un modelo de división por zonas y reconocida sin aprendizaje interactivo. Todo lo que tiene que hacer es iniciar el escaneado y guardar el texto reconocido, los pasos intermedios son llevados a cabo por Readiris.

READIRIS REPRODUCE EL DISEÑO DE SUS DOCUMENTOS

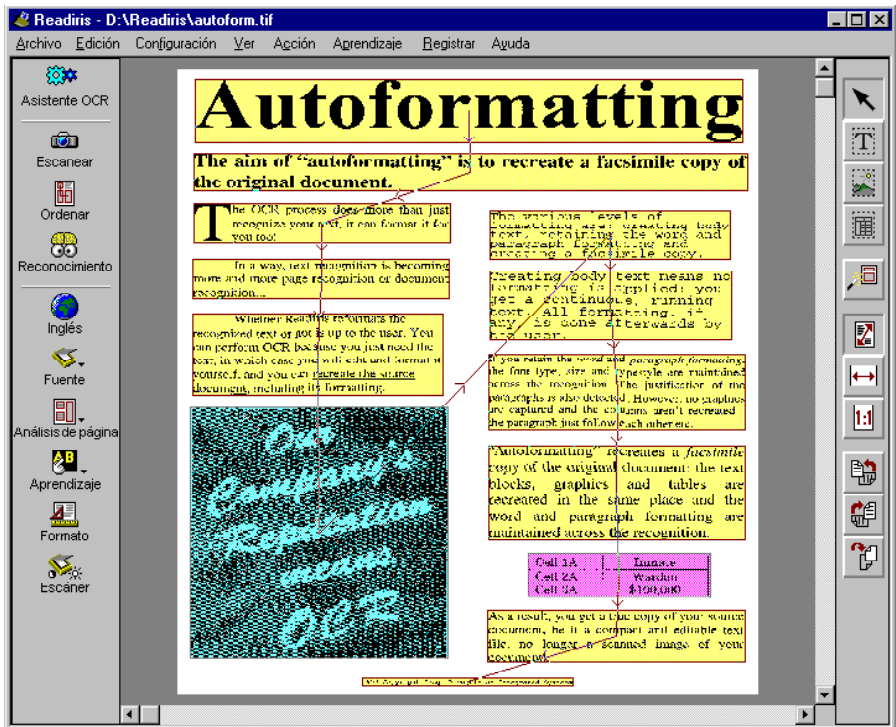
El reconocimiento automático que hace que el proceso de reconocimiento sea automático, no debe confundirse con el formato automático. “Formato automático” significa que Readiris reproduce una **copia facsímil** del documento escaneado, aplicando el formato de las palabras, párrafos y páginas del documento original.

Se utilizan tipos de letra iguales a los del documento original (serif y sans serif, proporcionales y fijos, normales y condensados); se mantienen los tamaños y estilos (negrita, cursiva y subrayado) durante el reconocimiento. Se reproducen las tabulaciones y la justificación (izquierda, centrada, derecha o total) de cada bloque de texto. La colocación de columnas, bloques de texto y gráficos imita la 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 “autoformato”; es decir, ¡puede incluir y excluir zonas, pero cualquier reclasificación no será tenida en cuenta!

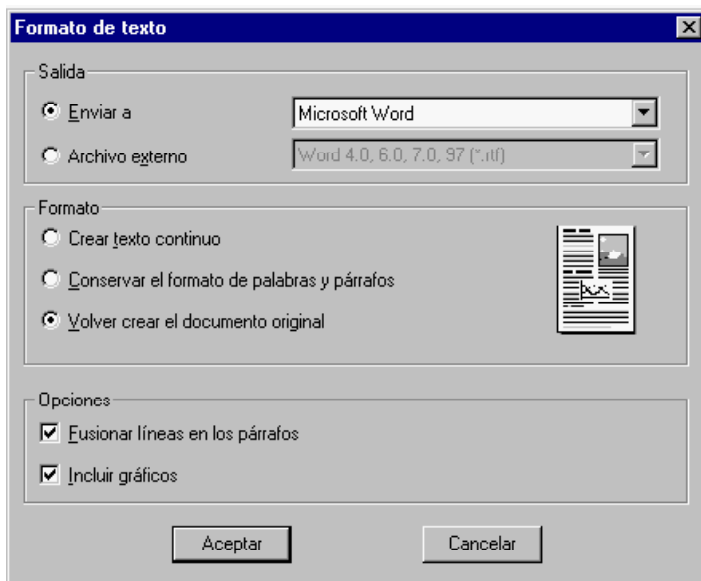
He aquí un ejemplo de cómo funciona. Para familiarizarse con esta función, abra el archivo gráfico AUTOFORM.TIF que se encuentra en la carpeta Readiris.



Haga clic en el botón “Formato” de la barra de herramientas principal y seleccione “enviar 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 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.)

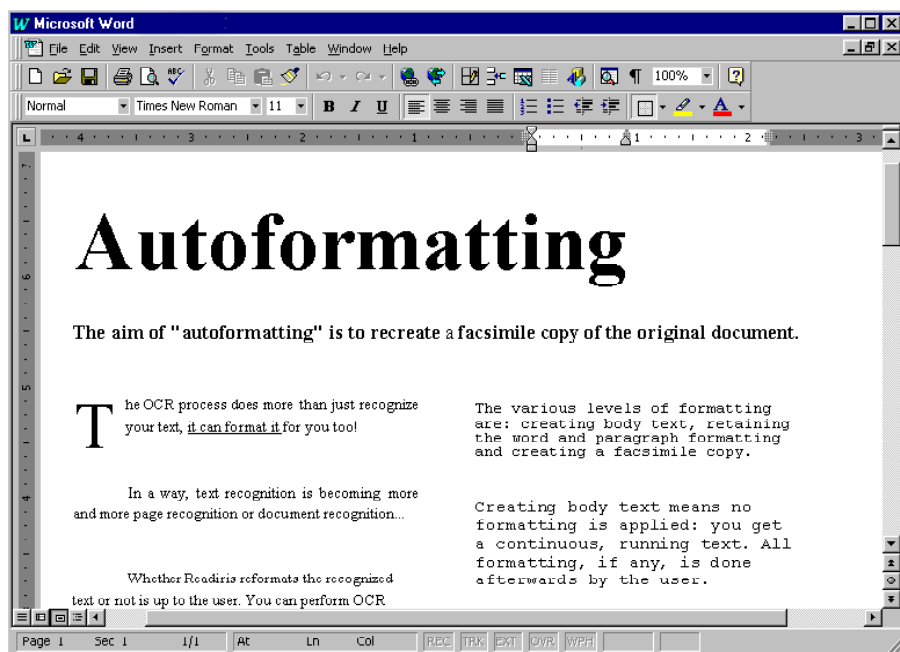


Observe que la reproducción del formato se limita a los tipos Word (DOC) y RTF, o bien a documentos enviados directamente a Microsoft Word 97. La ventaja es que Word (DOC) y RTF son formatos de texto muy extendidos, que pueden abrirse con casi cualquier procesador de textos.

Otros formatos “pobres” que generan texto “simple”, como ASCII, ANSI etc., *no* pueden utilizar los códigos avanzados de formato y, por consiguiente, no pueden ofrecer un formato automático. Del mismo modo, la aplicación WordPad es un editor de textos de características “reducidas” y no un procesador de textos avanzado; WordPad puede abrir archivos Word (DOC) y RTF, pero pasa por alto

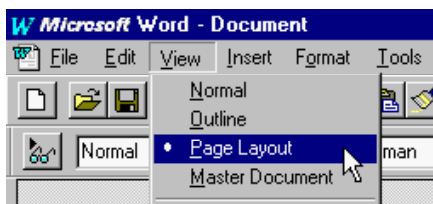
la mayoría de los códigos de formato, como los cuadros de texto, la justificación, etc.

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" se activa automáticamente!

Para ver este efecto correctamente, necesita activar el modo "WYSIWIG" de su tratamiento de textos, denominado generalmente modo "diseño de página".

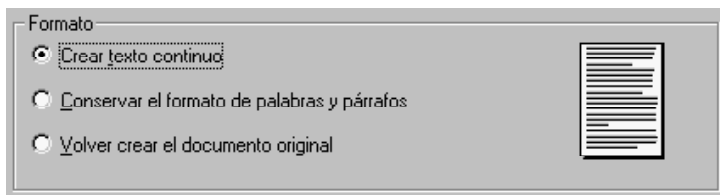


En pocas palabras, Readiris no sólo reconoce sus textos, sino que puede también formatearlos. Ahora, el OCR ya no es un simple reconocimiento de textos, sino que cada vez se acerca más a un **reconocimiento de páginas** y documentos.

FORMATO DE TEXTOS, 2ª PARTE

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

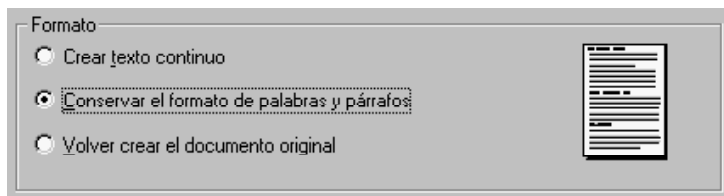
Como ilustra el icono de la derecha, crear **texto de cuerpo** significa crear un texto “continuo”, sin formato. El texto será capturado, pero se ignorará totalmente el formato. Utilice esta opción cuando necesite capturar un texto pero no su diseño.



(La generación de textos en lenguas asiáticas implica la creación de un texto independiente - es imposible aplicar la función de “autoformato” o copiar el formato de palabras o párrafos. Por otra parte, el número de formatos de archivo es reducido.)

La opción “Conservar el Formato de Palabras y Párrafos” representa un camino intermedio: durante el reconocimiento se mantiene el **formato de palabras** - tipo

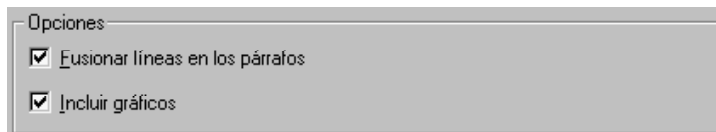
de letra (serif - sans serif, proporcional - fijo, normal - condensado), el tamaño de letra y el estilo (negrita, cursiva y subrayado)... Asimismo, también se mantiene el **formato de párrafo** - tabulaciones y justificación (izquierda, centrada, derecha o total).



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. Tampoco incluirá los gráficos como lo hace el formato automático completo.

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 sólo pueden guardarse con el formato automático “completo”, no con formatos de texto “pobres”, como ASCII, ANSI, etc.



Aún así, con Readiris puede guardar gráficos sin llevar a cabo el reconocimiento de texto. Dado que el *software* OCR de Readiris genera **imágenes en blanco y negro** y no imágenes en escala de grises o en color, no tiene sentido escanear fotografías con Readiris, pero sí puede capturar dibujos 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”.

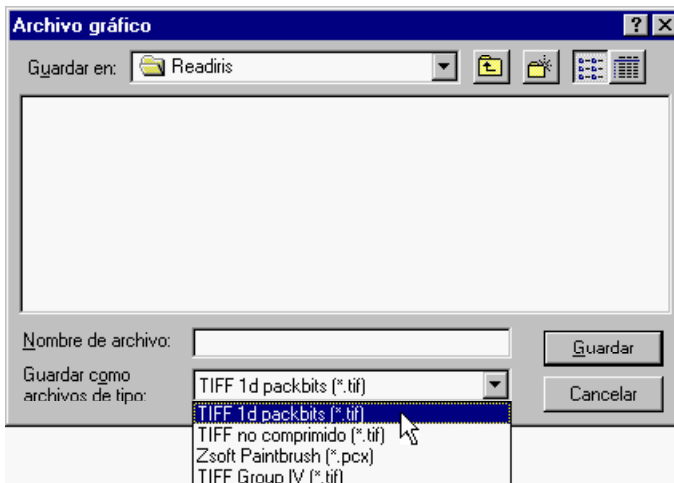


(Al igual que ocurre en los otros tipos de ventanas, la etiqueta emergente de la ventana de gráficos le informará de cuántas ventanas de este tipo existen. Por otro lado, si mantiene pulsada la tecla Alt mientras selecciona una ventana, podrá ver de qué tipo es y cuáles son sus coordenadas).

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 diseño y retoque. Dispone de los formatos TIFF y

Paintbrush (PCX). (Los archivos de Paintbrush siempre pueden abrirse con el accesorio de Windows denominado Paint.)



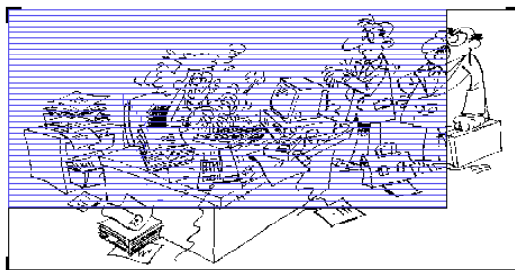
Los gráficos se guardan en un solo archivo. No tiene por qué limitarse a un solo gráfico, sino que puede trazar varias ventanas de gráfico y éstas serán “reunidas” en un solo archivo. (Puede utilizar la opción “Separar” de su programa de diseño gráfico para separarlas.)

No se permiten lados inferiores a 1 mm. – los mapas de bits de ese tamaño casi no contienen información. Las ventanas “irregulares” no rectangulares se convierten de forma invisible en rectángulos: Readiris cubre la máxima superficie



en función de las coordenadas más amplias. En el ejemplo a continuación, la ventana de gráfico con el borde negro se completa con la zona sombreada.

ur computer automatically.



nt is read by your scanner. This device acts as the "eye" of yo

LECTURA DE FAXES Y RECONOCIMIENTO DIFERIDO

Al guardar imágenes en archivos gráficos 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”. Se le preguntará en qué formato desea guardar la página: TIFF o Paintbrush (PCX).

Guardar toda la página como imagen...

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.

Readiris abre archivos TIFF en blanco y negro (sin comprimir, paquetes de bits y Grupo 3 y Grupo 4 comprimidos), Paintbrush (PCX) y mapas de bits de

Windows (BMP). Esta capacidad resulta especialmente útil para convertir sus **faxes** en archivos digitales de texto.

De ser posible, pida a sus corresponsales que envíen sus faxes con calidad “fina” – dichos faxes tienen una resolución superior (200 ppp) y darán mejores resultados al ser reconocidos.

No olvide que para abrir archivos gráficos puede **arrastrarlos con el ratón** desde el Explorador de Windows hacia la zona de imágenes de Readiris.

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 TABLE.TIF que se encuentra en la carpeta Readiris.

En realidad, la imagen TABLE.TIF 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.



Readiris - D:\Readiris\Table.tif

Archivo Edición Configuración Ver Acción Aprendizaje Registrar Ayuda

Asistente OCR

Escanear

Ordenar

Reconocimiento

Inglés

Huente

Análisis de página

Aprendizaje

Formato

Escáner

Reading Tables

Readiris recognizes tabular data and recreates them cell by cell in worksheets or as table objects inside wordprocessor files.

To insert tables as table objects, you must **retain** the word and paragraph formatting or recreate the source document; see the "Format" button on the main toolbar.

The page analysis detects "**gridded**" and "**ungridded**" tables. "Gridded" or "framed" tables have borders around the cells - as does the example below. The borders of the table cells get recreated.

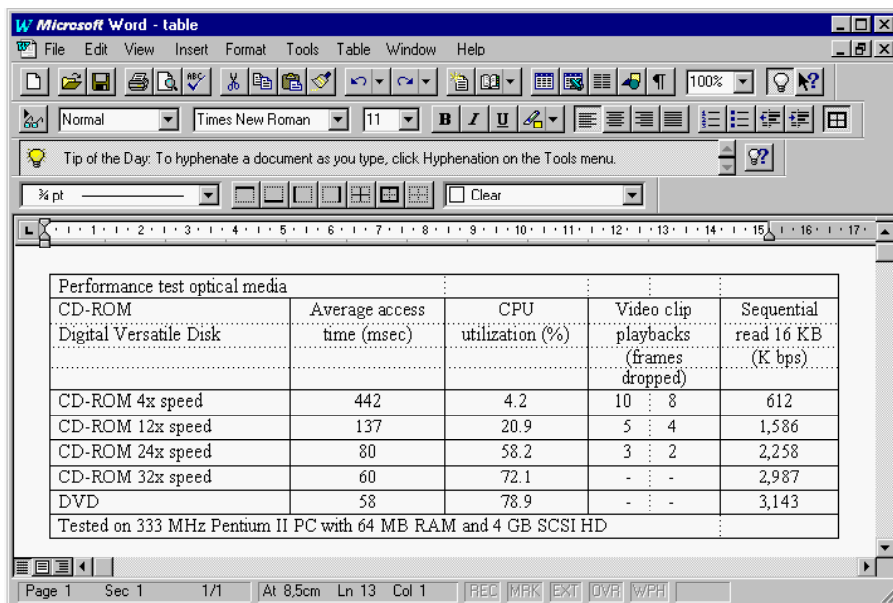
Performance test optical media				
CD-ROM Digital Versatile Disk	Average access time (msec)	CPU utilization (%)	Video clip playbacks (frames dropped)	Sequential read 16 KB (K bps)
CD-ROM 4x speed	442	4.2	10 8	612
CD-ROM 12x speed	137	20.9	5 4	1,586
CD-ROM 24x speed	80	58.2	3 2	2,288
CD-ROM 32x speed	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.

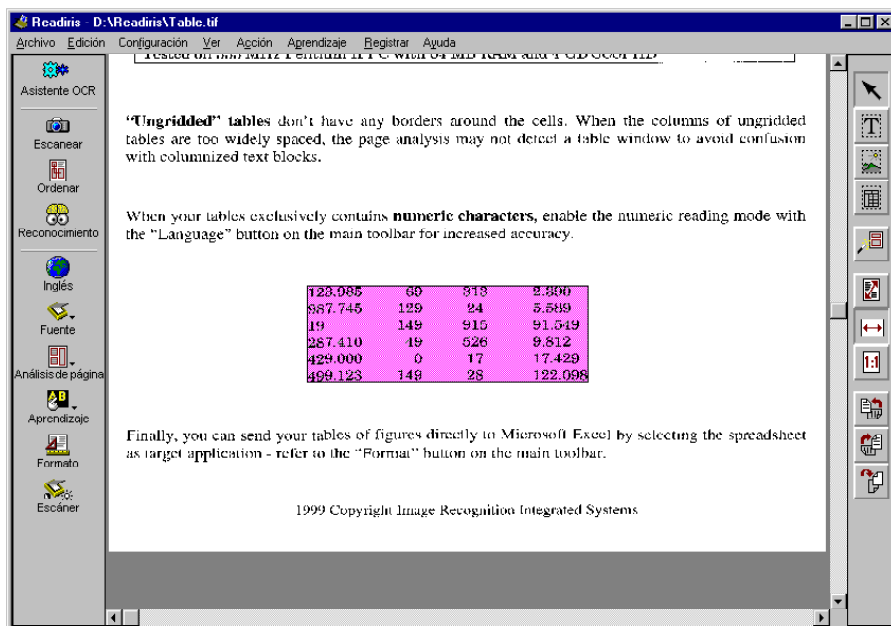
"**Ungridded**" tables don't have any borders around the cells. When the columns of ungridded tables are too widely spaced, the page analysis may not detect a table window to avoid confusion with columnized text blocks.

When your tables exclusively contains **numeric characters**, enable the numeric reading mode with the "Language" button on the main toolbar for increased accuracy.

Para obtener una reproducción exacta de la tabla, ejecute el reconocimiento con la opción "Conservar el Formato de Palabras y párrafos" o bien "Volver Crear el Documento Original". Abra un procesador de textos (si es que no ha enviado ya directamente la tabla reconocida a Word) y verá el resultado. (Naturalmente, también puede incluir párrafos de texto normales en este archivo.)



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!

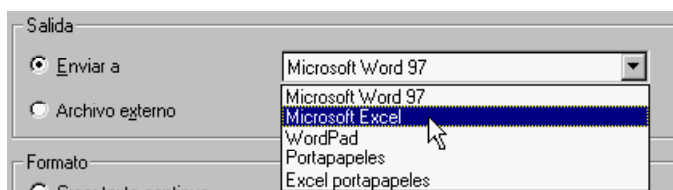


Para una precisión óptima de reconocimiento, limite el reconocimiento a los **símbolos numéricos** con el botón “Idioma”. (El modo numérico no es estrictamente numérico, incluye los símbolos 0 a 9, +, *, /, %, , (coma), . (punto), (,), -, =, \$ y £.)

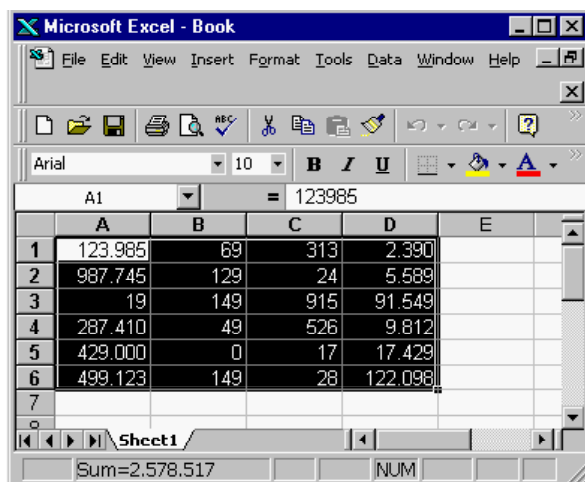


Puesto que esto sólo se puede hacer si la tabla no contiene símbolos alfabéticos – de lo contrario, las porciones de texto no se reconocerán correctamente – ahora podemos activar el modo numérico, algo que no podíamos hacer con la primera tabla.

En esta ocasión, enviaremos directamente el resultado del reconocimiento al programa de hoja de cálculo Microsoft Excel. Seleccionaremos, pues, Excel como aplicación de destino 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.



The screenshot shows the Microsoft Excel - Book window. The menu bar includes File, Edit, View, Insert, Format, Tools, Data, Window, and Help. The toolbar contains icons for file operations and editing. The spreadsheet has a formula bar showing 'A1 = 123985'. The data in the spreadsheet is as follows:

	A	B	C	D	E
1	123.985	69	313	2.390	
2	987.745	129	24	5.589	
3	19	149	915	91.549	
4	287.410	49	526	9.812	
5	429.000	0	17	17.429	
6	499.123	149	28	122.098	
7					

The status bar at the bottom shows 'Sum=2.578.517' and 'NUM'.

También puede enviar tablas al portapapeles - de ahí la opción “Enviar a - Portapapeles Excel” - o bien puede guardar dichas tablas en un archivo externo - y de ahí el formato de archivo Excel (el CSV es para datos separados por comas).



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 graficas y siga el procedimiento habitual; una vez más, la etiqueta emergente del botón le indicará el número de ventanas de tabla existentes.



LA FUNCIÓN “CONNECT”, OTRO ENLACE DIRECTO

Al igual que la función “Enviar a”, que exporta automáticamente el resultado de la lectura, la función “Connect” ofrece un vínculo directo de reconocimiento entre el escáner y las aplicaciones para Windows. La única diferencia es que el proceso se realiza en dirección contraria; es decir, el reconocimiento de los documentos no se efectúa en el programa Readiris para exportar posteriormente el resultado a una aplicación, ¡sino que se recurre a Readiris desde la propia aplicación de destino!

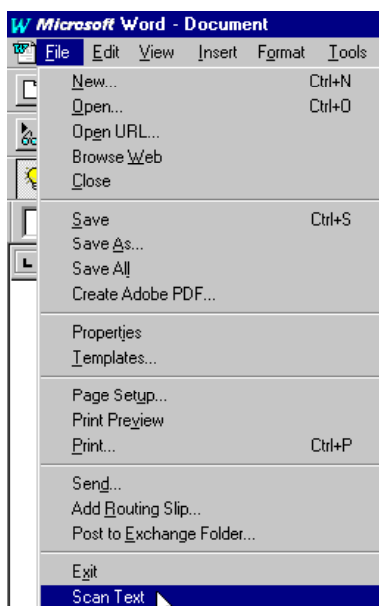
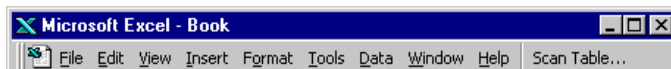
Ya que hay algunos pasos manuales, ningún procedimiento de **instalación** puede realizarla automáticamente, consulte el sistema de ayuda en línea o el archivo CONNREAD.DOC situado en la carpeta Readiris para aprender a instalarla – se requiere un macro para usar la función “Connect” desde Word, Excel y WordPerfect. (Este archivo “Léame” aparece automáticamente al final del programa de instalación.)



Correctamente instalada, la función "Connect" añade una nueva opción a los menús de su aplicación. En los ejemplos a continuación se añade la opción



“Escanear Texto” en el menú “Archivo” de Word y el menú “Escanear Tabla” en la barra de menús de Excel.



He aquí como funciona: Readiris *no* puede estar abierto al accionar el macro “Connect”. Pulse la nueva opción en el menú de su aplicación para ejecutar Readiris. El interfaz “Connect” puede hacerle algunas preguntas antes de que Readiris empiece a funcionar. Proceda de forma habitual y salga del interfaz “Connect” cuando haya terminado. Los resultados del reconocimiento se insertan en la posición del cursor.

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.

