

Background

package

READ IRIS OCR

yananár

real

text

Save Time, No More Retyping!

typed

IRIS

Druckqualität

trainable

paper

učici modul

OCR

обраа

matrix

caractères accentués

omnifont

easy

recognition

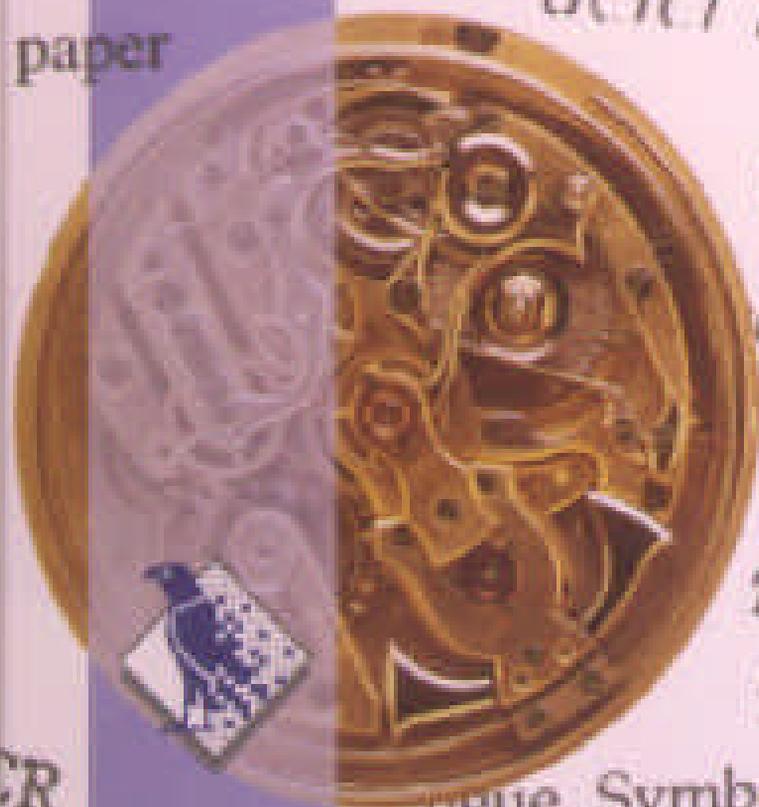
fax

neue Symbole und Schriftar

IR

IMAGE RECOGNITION
INTEGRATED
SYSTEMS

JUST READ IT!



Readiris é uma marca comercial da Image Recognition Integrated Systems.

As informações contidas neste documento estão sujeitas a alterações sem aviso prévio e não representam um compromisso. O software descrito neste documento está sujeito a um contrato de licenciamento ou de sigilo. O software somente poderá ser usado ou copiado de acordo com os termos do contrato.

Proximity é uma marca comercial registrada da Proximity Technology Inc.

IBM é uma marca comercial registrada da International Business Machines Corp. DOS e Windows são marcas comerciais (registradas) da Microsoft Corp. Todos os modelos de digitalizadores e de formatos de arquivo mencionados neste guia são marcas comerciais (registradas) dos respectivos proprietários.

THE PROXIMITY/MERRIAM-WEBSTER LINGUIBASE

Copyright 1984 Todos os direitos reservados

Proximity Technology Inc.

Copyright 1984 Todos os direitos reservados

Merriam-Webster Inc.

THE PROXIMITY/MERRIAM-WEBSTER LINGUIBASE

Copyright 1984 Todos os direitos reservados

Proximity Technology Inc.

Copyright 1984 Todos os direitos reservados

William Collins Sons & Co. Ltd.

A ortografia deste produto se baseia na Proximity Linguistic Technology

Copyright 1994 Image Recognition Integrated Systems

Todos os direitos reservados

BEM-VINDO

O Readiris é um poderoso sistema de Reconhecimento Ótico Inteligente de Caracteres (IOCR), que transforma o computador e o digitalizador em um eficiente sistema de leitura de páginas.

É o modo melhor e mais rápido de transformar informações impressas e dactilografadas em arquivos de texto que podem ser editados e usados pelo processador de texto, planilha, software de editoração eletrônica, etc.

O Readiris usa menus conhecidos e o estilo do ambiente Windows. O Readiris é tão simples que até pode ser executado a partir de seus aplicativos favoritos do Windows!



SOBRE O READIRIS

O Readiris é um programa muito fácil de instalar e usar. Você precisa dois minutos para a instalação e cinco para começar a trabalhar. Na maioria das vezes, tudo o que você precisa fazer é clicar a barra de botões.

Embora projetado como um produto de baixo custo, o Readiris lhe permite acessar as técnicas mais avançadas da inteligência artificial usadas no campo IOCR (Reconhecimento Ótico Inteligente de Caracteres).

O Readiris é um sistema de aprendizagem automática. Por outras palavras, você pode aprender novos caracteres. Vários bancos de dados lingüísticos e léxicos são usados para analisar o contexto lingüístico e fornecer uma poderosa ferramenta para encontrar novas soluções em casos difíceis.

Finalmente, o Readiris permite a você capturar imagens de documentos impressos através do digitalizador e as importar para seu aplicativo de gráficos.



SOBRE ESTE GUIA

Este guia descreve como usar o Readiris. O primeiro capítulo, “Como começar”, fornece informações sobre os requisitos do sistema e sobre os procedimentos de instalação.

O segundo capítulo “Começo rápido”, fornece uma panorâmica geral sobre o funcionamento do Readiris e informações sobre conceitos básicos.

Após esse breve tutorial, o terceiro capítulo, “Aprendendo mais sobre o Readiris”, fornece mais detalhes sobre várias funções avançadas do Readiris.

O quarto capítulo, “Referência de comandos”, fornece uma descrição completa dos menus e respectivos comandos do Readiris.

O último capítulo, “Mais sobre OCR”, fornece uma breve descrição dos conceitos e parâmetros do OCR. Descreve o impacto desses diferentes parâmetros no desempenho dos sistemas OCR em geral e no Readiris em particular.

Toda a informação complementar não incluída neste guia poderá ser encontrada no arquivo README.TXT. Leia este arquivo se tiver problemas com a instalação ou funcionamento do Readiris. Para maiores informações sobre digitalizadores, consulte o arquivo READSCAN.TXT.



ÍNDICE

Bem-vindo.....	3
Sobre o Readiris	5
Sobre este guia	7
Índice	9

1. Como começar

Requisitos do sistema	13
Instalando o Readiris.....	13
Instalando a capacidade Connect (conexão).....	14

2. Começo rápido

Executando o Readiris	15
Definindo sua configuração padrão.....	16
OCR em uma imagem previamente varrida.....	17
Definições	17
Abrindo o arquivo de imagens	17
Zoom da imagem	19
Definindo a janela Text e Graphic	19
Modificando uma janela	20
Reconhecimento de caracteres	20
Vendo os resultados	22
OCR em um documento impresso.....	22
Configurando o digitalizador	22
Varrendo com um digitalizador plano	23
Varrendo com um digitalizador portátil	23
Executando o reconhecimento	24

3. Aprendendo mais sobre o Readiris

Definindo configurações padrão	25
Configurando o digitalizador plano	26
Configurando o digitalizador portátil	27
Formato de texto	27
Formato de gráficos	28
Idioma	29
Salvando predefinições	30
Salvando e restaurando definições.....	30
Salvando definições	30



Restaurando definições	31
Tudo sobre o Painel de Controle.....	31
Document	32
Process	33
Files	35
Varrendo uma página com um digitalizador plano.....	37
Varrendo uma página com um digitalizador portátil.....	37
Olhando de perto a imagem varrida.....	38
Controlando a qualidade da imagem varrida.....	39
Ferramentas de representação em janela.....	40
Criando uma janela retangular	40
Criando uma janela poligonal	41
Selecionando janelas	41
Modificando uma janela	42
Movendo janelas	42
Eliminando janelas	42
Reconhecimento de caracteres	43
Aprendendo sobre fontes	44
OK	45
Continue	46
Confirm	46
Undo	46
Delete	46
End	46
Usando corretamente os dicionários de fontes.....	47
Ajuda	48
Reconhecendo tabelas.....	48
Reconhecimento automático	50
Usando a capacidade “Connect” (conexão).....	51
WordPerfect, Word e Excel	51
Outros aplicativos do Windows	52

4. Referência a comandos

Menu File	55
Open	55
Get Info	55
Select Source	55
Acquire	55
Save Layout	56
Get Layout	56
Save Settings	56

Get Settings	56
Set Default Settings	56
Exit	56
Menu Edit	56
Undo	57
Delete	57
Copy	57
Paste	57
Clear	57
Select All	57
Menu Preferences	57
Scanner	58
Text Format	58
Graphics Format	58
Language	58
Toolbar	58
Menu View	58
Fit in Window	59
50% Actual Size	59
Actual Size	59
200% Actual Size	59
All Windows	59
Selected Type	59
Sort	59
Sort View	59
Sort Reset	60
Rotate	60
Toolbox	60
Coordinates	60
Menu Process	60
Control Panel	60
OCR	61
Interactive Learning	61
Page Analysis	61
View after Scan	61
Control Panel	61
Pitch	62
Size	62
Font	62
Pages(s) to Read	62
Page Analysis	62
View after Scan	62
Interactive Learning	63
Strong Checking	63



Input	63
Output	63
Layout	63
Dico	64
Scan	64
Close	64
Menu Help	64
Contents	64
About Readiris	64

5. Mais sobre OCR

Processo de reconhecimento ótico (OCR).....	65
Âmbito do Readiris	67
Número de caracteres e contexto lingüístico	67
Isolamento de caracteres	68
Gama de tamanho de caracteres	68
Documentos excluídos	68

1. COMO COMEÇAR

Este capítulo mostra como você deve instalar o Readiris antes de o utilizar pela primeira vez.

Este guia assume que você já está familiarizado com seu computador e com o ambiente do Windows.

REQUISITOS DO SISTEMA

Estes são os requisitos mínimos para usar o Readiris:

- Computador IBM 386, PS/2 ou compatível. Multitarefa suportada mesmo em um computador 386!
- 4 MB de RAM. Você pode precisar espaço de memória adicional para usar a capacidade "Connect" (conexão).
- Sistema operativo MS-DOS 3.3 ou versão posterior. Recomenda-se o MS-DOS 6.x.
- Licença do Windows 3.1. Não execute o Windows 3.1. no modo standard!

INSTALANDO O READIRIS

Como medida de segurança, faça uma cópia de reserva dos disquetes do Readiris e os guarde em um local seguro.

Efetue os seguintes passos para instalar o Readiris:

- Certifique-se de estar no Program Manager (Gerenciador de Programas) do Windows. Selecione Run no menu File e escreva **a:setup** na caixa da linha de comando e clique OK.
- O programa de instalação é executado. Indique o idioma pretendido e siga as instruções que são exibidas na tela.



Então o sistema pede para você inserir o segundo disquete.

O grupo do programa e os ícones do aplicativo são criados pelo programa de instalação.

INSTALANDO A CAPACIDADE CONNECT (CONEXÃO)

O Readiris deverá ser corretamente instalado antes de você instalar a capacidade de conexão.

Execute o Readiris, selecione o digitalizador e salve as predefinições antes de instalar a capacidade Connect. Para mais informações, consulte o capítulo 3, “Aprendendo mais sobre o Readiris”.

Para o WordPerfect, Word e Excel, você precisa instalar uma macro se pretender usar a capacidade de conexão. O sistema de ajuda on-line lhe dirá como fazer - o tópico de ajuda apropriado é exibido automaticamente quando você executa o programa de instalação do Readiris.

Para todos os outros aplicativos do Windows, o programa de instalação do Readiris instalou um servidor de conexão de tarefas; o usuário não é envolvido.

2. COMEÇO RÁPIDO

A melhor maneira de ver como o Readiris funciona é usando-o. As páginas seguintes lhe permitem começar imediatamente. Juntamente com a licença do programa é fornecida uma imagem previamente varrida, BRAZIL.TIF, para que você comece treinando no programa OCR, mesmo que o digitalizador ainda não esteja ligado.

EXECUTANDO O READIRIS

Certifique-se de estar no Program Manager (Gerenciador de Programas) do Windows. Clique duas vezes o ícone Readiris.



A tela de inicialização é exibida.

A janela do aplicativo Readiris contém não só menus de comandos como também uma **barra de botões**. Basta clicar esses botões para executar a maior parte das ações. Clique o comando Toolbar no menu Preferences para ativar ou desativar a barra de botões.

DEFININDO SUA CONFIGURAÇÃO PADRÃO

1. Selecione o menu Preferences. São exibidos os seguintes comandos: Scanner, Text Format, Graphics Format e Language.
2. Em cada um desses comandos selecione as predefinições.
3. Selecione o comando Set Default Settings no menu File. Clique OK para confirmar.

Pode então usar o sistema. Ainda que você tenha um digitalizador ligado, recomendamos que experimente primeiro o tutorial.

OCR EM UMA IMAGEM PREVIAMENTE VARRIDA

Este breve tutorial usa uma imagem previamente varrida de um texto sobre OCR. Vamos reconhecer o texto e salvar a arte-final em um arquivo gráfico.

Definições

1. Certifique-se de ter selecionado o idioma Brazilian na lista de cortina Language da barra de botões.



2. No menu Preferences, selecione o formato ASCII para o texto com o comando Text Format e o formato Paintbrush (PCX) para o arquivo gráfico com o comando Graphics Format.
3. Escolha Control Panel no menu Process.
4. Cheque se o tamanho do caractere e o tipo do fonte estão definidos para normal e se o espaçamento dos caracteres está definido para proporcional.
5. Verifique se o dicionário dos fontes é o READIRIS.DUS e se está no modo New.
6. Se necessário, desative Page Analysis na barra de botões.
7. Feche o Control Panel clicando o botão Close.

Abrindo o arquivo de imagens

1. Selecione a opção View After Scan no menu Process.
Essa opção é selecionada se tiver uma marca de verificação.
Caso contrário, clique View After Scan.

Zoom da imagem

Você pode ampliar partes da imagem.

1. Coloque o ponteiro do mouse na área da imagem que pretende ampliar.
2. Clique o botão *direito* do mouse para ampliar. Clicando novamente esse botão restaura o tamanho total da imagem.

Você também pode usar os comandos do menu View e as teclas de atalho associadas. Essas teclas podem mesmo ser usadas durante o movimento do mouse, o que é especialmente útil quando você estiver trabalhando em páginas com um layout complexo.

Definindo a janela Text e Graphic

Na segunda fase, você indica os blocos de texto que considera importantes. Vamos desenhar uma janela de texto à roda do texto a reconhecer e uma janela gráfica à roda da arte-final.

1. Selecione a ferramenta do retângulo, clicando a ferramenta retangular na caixa de ferramentas Windows. O ponteiro do mouse se transforma em uma cruz.



2. Selecione o tipo de janela: texto.
3. Mova a cruz para perto do canto superior esquerdo do texto e clique.
4. Aumente a janela movendo a cruz para o canto inferior direito e clique novamente.

A janela de texto está desenhada.

Para desenhar uma janela gráfica, proceda de forma idêntica, sendo a única diferença você precisar selecionar o tipo de janela gráfica na caixa de ferramentas Windows.



Modificando uma janela

1. Selecione a janela clicando a ferramenta da seta, na caixa de ferramentas Windows.



2. Mova o cursor para um dos marcadores.
3. Clique e segure o botão esquerdo do mouse.
4. Mova o mouse. A janela é ajustada de acordo com o movimento do mouse.
5. Libere o botão quando a janela atingir o tamanho correto.

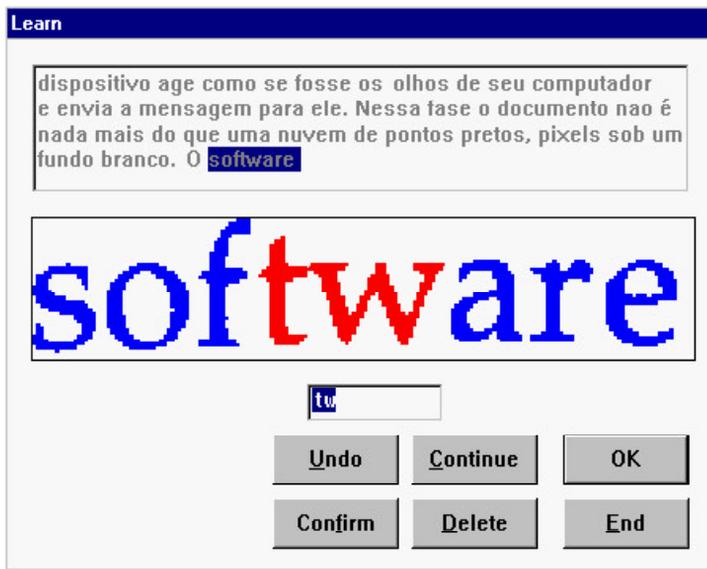
Reconhecimento de caracteres

Quando suas janelas estiverem corretamente definidas, você pode efetuar o processo de reconhecimento de caracteres.

1. Clique o botão OCR na barra de botões.



No final da fase de reconhecimento, o programa entra no modo de aprendizagem interativa. O texto reconhecido é exibido progressivamente e o sistema pára nos caracteres duvidosos. O sistema exibe então a imagem de cada um dos caracteres duvidosos integrados no respectivo contexto e propõe uma solução.



2. Clique OK ou pressione Enter para confirmar uma proposta correta; escreva a solução correta se a proposta do sistema está errada e então clique OK ou pressione Enter. O botão Continue permite a você corrigir sem a aprendizagem. Por enquanto ignore os outros botões.

(De fato, tem muito mais do que isso. O significado exato desses botões será explicado no capítulo seguinte.)

3. Quando terminar, saia do Readiris.

Uma vez que o botão OCR também pode ser usado para salvar gráficos, a zona de gráficos foi automaticamente salva.



Vendo os resultados

O arquivo de texto READIRIS.TXT e o arquivo gráfico READIRIS.PCX devem estar em seu diretório de trabalho.

1. Execute o Notepad/WordPad e abra o arquivo READIRIS.TXT. O texto reconhecido é exibido na tela. O documento impresso se transformou em um arquivo de texto editável!
2. Saia do Notepad/WordPad.
3. Execute o Paint/Paintbrush e abra o arquivo READIRIS.PCX. O módulo gráfico do Windows exibe a imagem. Agora você pode editar essa figura.

OCR EM UM DOCUMENTO IMPRESSO

Agora é assumido que você instalou corretamente o digitalizador.

Configurando o digitalizador

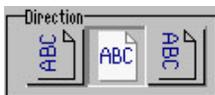
1. Clique o botão Scanner na barra de botões. Os digitalizadores suportados são listados. Selecione o modelo do digitalizador instalado e clique OK. Os formatos de página suportados são listados para os digitalizadores planos, bem como as várias resoluções.



2. Para fazer o OCR, selecione uma resolução de 300 dpi. Se a imagem é de má qualidade, pode usar 400 dpi.
3. Você pode precisar ajustar a definição do brilho e do contraste com a barra de rolagem, se disponível. Nos digitalizadores portáteis, o brilho e a resolução são definidos no próprio digitalizador.
4. Indique a fonte de entrada correta clicando Scanner na barra de botões.

Varrendo com um digitalizador plano

1. Coloque o documento na mesa do digitalizador.
2. Selecione a orientação de varredura correta na barra de botões.



3. Clique o botão Scan.

A página é varrida e exibida progressivamente na tela.

Se você selecionou a orientação de varredura Landscape, a imagem é rodada imediatamente após a varredura.

Varrendo com um digitalizador portátil

1. Selecione a direção de varredura correta na barra de botões.
2. Clique o botão Scan. O cursor se transforma em um pequeno digitalizador portátil.
3. Coloque o digitalizador na seção do documento que pretende capturar. O documento deverá estar sobre uma superfície plana.
4. Carregue no botão Start do digitalizador, em o movendo na direção selecionada.

O movimento terá de ser lento e tão regular quanto possível, caso contrário a imagem varrida poderá ficar inutilizada para o OCR. A imagem é exibida progressivamente, à medida que o digitalizador se move.

5. Após terminar a varredura da imagem, pressione a barra de espaço para passar o controle para o Readiris.

Se selecionou a direção de varredura da esquerda para a direita ou da direita para a esquerda, pressione a barra de espaço para rodar diretamente a imagem.



Executando o reconhecimento

Pode usar o Readiris que nem você fez com a imagem previamente varrida: pode ampliar para ver com maior detalhe, indicar as zonas de interesse desenhando janelas, executar o reconhecimento, etc.

3. APRENDENDO MAIS SOBRE O READIRIS

Este capítulo fornece vários detalhes sobre as capacidades do Readiris.

Você irá aprender como varrer uma imagem e como controlar a qualidade da imagem varrida. Após efetuar a varredura do documento, poderá definir janelas para especificar quais as partes do texto que você pretende sejam reconhecidas pelo Readiris ou fazer o sistema analisar automaticamente o layout de suas páginas. Então o Readiris inicia o reconhecimento de caracteres e gera um arquivo de saída.

Vamos ver igualmente algumas funções avançadas, como o reconhecimento automático, a capacidade “Connect” (conexão) e vamos aprender como fazer o OCR de tabelas.

DEFININDO CONFIGURAÇÕES PADRÃO

Primeiro, você precisa definir suas configurações, ou seja, deverá especificar:

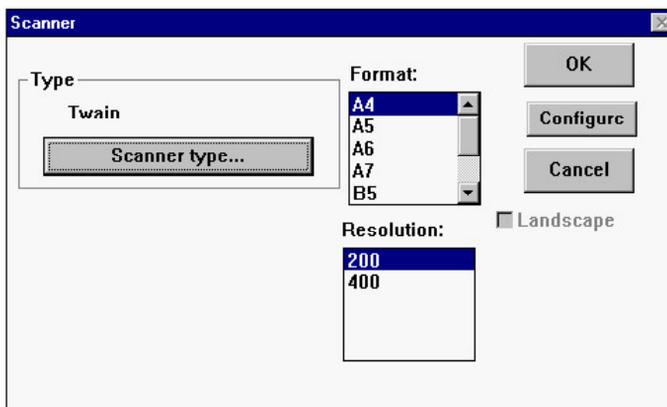
- o digitalizador suportado e suas definições
- o formato do arquivo para o texto e gráficos de saída
- o idioma do texto a ser reconhecido
- os parâmetros definidos com o painel de controle

Essas definições raramente precisam ser alteradas e poderão ser salvas em um arquivo de predefinições, evitando assim que você precise as definir novamente.



Configurando o digitalizador plano

1. Clique o botão Scanner na barra de botões. Os digitalizadores suportados são exibidos.



2. Selecione o digitalizador plano instalado e clique OK. É exibida a resolução predefinida, bem como o formato de página, o brilho e possivelmente o contraste.
3. Clique o formato de documento correspondente às páginas que você geralmente lê. A4 é o formato de página predefinido.
4. Selecione a resolução de varredura. A resolução predefinida é 300 dpi.
5. Defina o brilho e o contraste do digitalizador. A posição central é o valor padrão.

A opção de contraste poderá não estar disponível em seu digitalizador. As barras de rolagem de brilho e de contraste permitem a você ajustar a varredura para documentos mais claros ou mais escuros. Os digitalizadores HP permitem ativar a opção AutoExp, que ajusta automaticamente o brilho da varredura de acordo com o documento.

O Readiris também suporta a tecnologia **AccuPage** HP. Esta tecnologia permite ajustar o brilho do digitalizador autonomamente, bem como usar as técnicas de processamento de imagens para otimizar a imagem varrida para o processo de reconhecimento.

Esta tecnologia se encontra disponível somente nos digitalizadores HP. Selecione o “scanner” AccuPage em vez de um digitalizador HP e clique a opção AccuPage.

A documentação técnica fornecida com o digitalizador lhe fornece maiores informações sobre a respectiva configuração e funcionamento.

O arquivo READSCAN.TXT do diretório Readiris contém igualmente informação sobre como configurar um digitalizador.

Configurando o digitalizador portátil

1. Clique o botão Scanner na barra de botões. Os digitalizadores suportados são exibidos.
2. Selecione o digitalizador portátil instalado e clique OK. A resolução atual é exibida na tela.

A resolução e o brilho são definidos no digitalizador.

A documentação técnica fornecida com o digitalizador portátil lhe fornece maiores informações sobre sua configuração e funcionamento.

O arquivo READSCAN.TXT do diretório Readiris também contém informação sobre como configurar um digitalizador.

Formato de texto

1. Selecione Text Format no menu Preferences.



2. Selecione o formato de texto pretendido.

O Readiris sai diretamente para vários **processadores e formatos de texto**. Use o formato ASCII ou ANSI para outros processadores de texto.

O Readiris também envia diretamente tabelas para **planilhas** em um formato de tabela genérico.

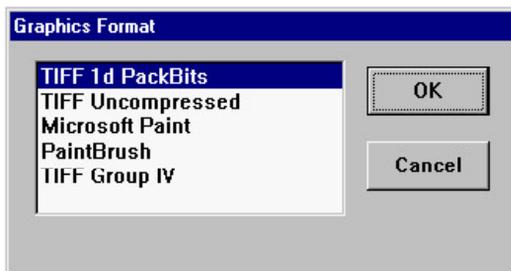
O “formato” da **área de transferência** lhe permite salvar os resultados diretamente na área de transferência e colar diretamente no aplicativo do Windows.

As duas opções seguintes permitem influenciar a aparência do resultado:

- Parágrafo** (Paragraph): ativa a detecção de parágrafos. Quando ativado, o Readiris respeita a mudança automática de linha, típica dos processadores de texto, colocando um retorno de carro somente no final dos parágrafos.
- Edentação** (Indentation): ativa os códigos da margem do texto. Isto é útil se você pretende preservar o aspecto do texto original. Caso deseje importar seu texto em um aplicativo de editoração eletrônica, não ative esta função!

Formato de gráficos

1. Selecione Graphics Format no menu Preferences.
2. Selecione o formato de gráficos aceito pelo aplicativo de gráficos.



O Readiris salvará seus gráficos nesse formato gráfico.

Idioma

É importante que o usuário indique o idioma do documento a ser reconhecido para que os **bancos de dados lingüísticos** desempenhem um papel positivo no processo de reconhecimento.

1. Clique a lista de cortina na barra de botões ou selecione o comando Language no menu Preferences.
2. Selecione o idioma que pretende reconhecer mais freqüentemente.



O Readiris está equipado com vários módulos lingüísticos que influenciam positivamente os níveis de reconhecimento!

O Readiris também poderá ser definido para um **modo numérico** a fim de reconhecer tabelas de números com a máxima precisão e velocidade.



Salvando predefinições

Você pode armazenar essas definições para evitar as definir novamente toda vez que executar o Readiris.

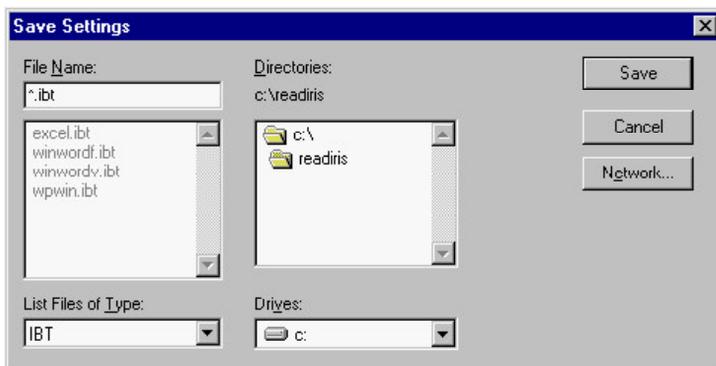
Selecione Save Default Settings no menu File. Estas definições serão carregadas toda vez que executar o Readiris. Pode ver essas definições com os comandos Get Info do menu File.

SALVANDO E RESTAURANDO DEFINIÇÕES

Adicionalmente, você poderá salvar definições específicas para outros aplicativos que não o aplicativo básico. Essas definições são salvas com o comando Save Settings e ativadas com o comando Get Settings do menu File.

Salvando definições

1. Selecione Save Settings no menu File. Os arquivos de definições do diretório atual são exibidos.
2. Escreva um nome de arquivo para salvar suas definições em um arquivo novo ou para sobrescrever um arquivo existente.



3. Clique Save.

Este procedimento armazena a configuração atual do Readiris no arquivo especificado. Esses arquivos têm a extensão predefinida *.ibt.

Restaurando definições

1. Selecione Get Settings no menu File.
Os arquivos de definições do diretório atual são exibidos.
2. Selecione o arquivo correto ou selecione um arquivo de outro diretório.
3. Clique Get.

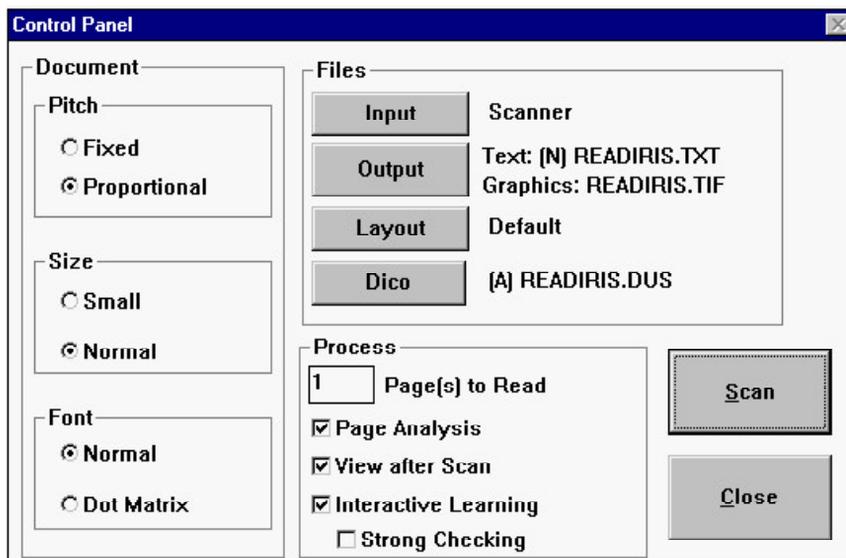
TUDO SOBRE O PAINEL DE CONTROLE

Antes de começar o processo OCR em um documento, você precisa especificar as características desse documento, ou seja:

- o idioma do documento
- a qualidade do documento
- o tipo do fonte
- o tamanho do fonte e a densidade horizontal

Você também pode especificar se existe um dicionário de fontes para os tipos de letra do documento a ser varrido.

O Control Panel lhe permite controlar todos os parâmetros necessários para ajustar o software a vários tipos de documentos que podem ser reconhecidos.



O Control Panel está dividido em três partes:

- Document
- Process
- Files

Document

Esta parte especifica as seguintes características do documento a ser reconhecido:

- a densidade horizontal (pitch): proporcional ou fixa
- o tamanho dos caracteres (size): normal ou pequeno
- o tipo do fonte (font): normal ou matriz de pontos

Com uma **densidade horizontal fixa**, todos os símbolos do fonte têm a mesma largura, por exemplo, na maioria dos documentos dactilografados.

Uma **densidade horizontal proporcional** significa que o tamanho de um caractere depende da sua forma. Um "m" será mais largo que um "i". A maioria dos livros e jornais são impressos com densidade horizontal proporcional.

O tamanho é considerado **pequeno** quando o tamanho do ponto é menor que 10 e quando o espaçamento entre linhas é demasiado pequeno. Nesse último caso, o Readiris teria problemas para separar corretamente duas linhas consecutivas.

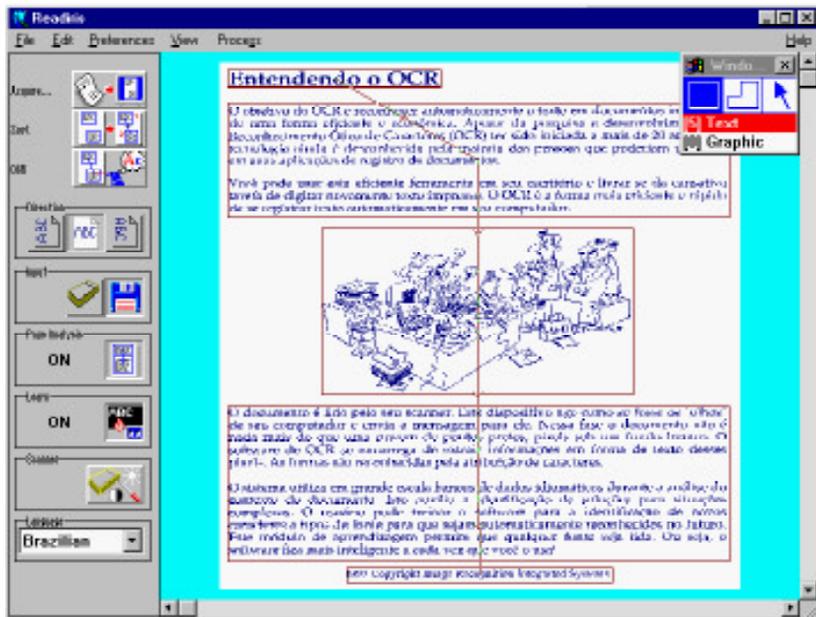
O usuário também deverá indicar o tipo de fonte: você não pode fazer o reconhecimento ótico de caracteres de **matriz de pontos** sem ativar o parâmetro correto. A qualidade dos caracteres de matriz de pontos não é a melhor e, muitas vezes, estão cortados ou são compostos por pontos individuais. Veja o exemplo abaixo.

ape--descended life

Process

Define o modo de processamento do Readiris.

Quando a opção **Page Analysis** é ativada, os blocos de texto e de gráficos da página são detectados e classificados automaticamente. Isso evita ao usuário desenhar manualmente uma janela à roda da imagem lida. Esse processo é bastante útil quando você faz o reconhecimento ótico de textos por colunas e de documentos com um layout complexo.



Inicialmente, a análise de página considera todos os blocos detectados como sendo blocos de texto. Não interfira quando aparecerem gráficos no documento: em uma segunda fase, o software detecta automaticamente a inexistência de caracteres nesse ponto do documento! (Também não serão salvados como gráficos.)

Os blocos são classificados de cima para baixo e da esquerda para a direita, de forma a abranger documentos com colunas, mas você poderá alterar a ordem da classificação. Clique o botão Sort na barra de botões e um marcador de janela das várias janelas. As janelas não clicadas neste ponto serão simplesmente ignoradas!



específico. Os layouts são configurações de representação em janela e contêm a estrutura das janelas definida durante a prévia de documentos. Após a representação em janelas manual e a análise de páginas pelo sistema, os layouts são o terceiro método para indicar zonas de interesse.

Dico: define o dicionário de fontes e o modo.

Estes fontes ou dicionários do usuário contêm informações sobre os caracteres com os quais você treina no sistema durante a fase de validação.

Existem três modos de dicionários disponíveis:

- New:** você cria um novo dicionário de fontes, começando sem quaisquer conhecimentos anteriores.
- Append:** você usa os conhecimentos sobre fontes adquiridos anteriormente, em os completando.
- Read:** você usa os conhecimentos adquiridos anteriormente mas não pode armazenar novas informações.

É muito importante usar os dicionários de fontes de forma correta para obter um reconhecimento rápido e preciso de seus documentos. Para tal, você precisa selecionar o modo correto para os dicionários.

Caso deseje reconhecer somente uma página, selecione o modo New. Entretanto, caso pretenda reconhecer várias páginas do mesmo tipo - ou seja, documentos com o mesmo tipo de letra e qualidade de impressão - você deve selecionar o modo New para a primeira página, o modo Append para a segunda e terceira página e o modo Read para as restantes páginas.

Salvar sempre para o mesmo dicionário do usuário não faz sentido: o tamanho destes arquivos é limitado!

VARRENDO UMA PÁGINA COM UM DIGITALIZADOR PLANO

1. Selecione a orientação de varredura adequada na barra de botões. (Você também pode ativar a opção Landscape clicando o botão Scanner na barra de botões.)
2. Clique o botão Scan.

O documento é varrido imediatamente. Quando a opção Preview é ativada, a imagem é exibida progressivamente na tela.

Se você selecionou Landscape como orientação da varredura, a imagem é rodada diretamente. Também poderá rodar as imagens em seguida com o comando Rotate do menu View.

Preste atenção ao **desalinhamento**. Embora a análise de página e o reconhecimento efetuado pelo Readiris tolerem desvios, pode ficar difícil representar uma janela e fazer o reconhecimento ótico correto de uma página varrida se o desalinhamento for demasiado significativo.

Clique o botão Scanner para alterar as definições do digitalizador, por exemplo, a resolução, o formato de página, etc.

VARRENDO UMA PÁGINA COM UM DIGITALIZADOR PORTÁTIL

Após ativar todos os parâmetros necessários, você poderá ler seu documento.

1. Selecione a orientação de varredura adequada na barra de botões.
2. Clique o botão Scan na barra de botões. O cursor se transforma em um pequeno digitalizador portátil.
3. O brilho e a resolução são definidos no próprio digitalizador. Use uma resolução de 300 dpi para documentos de boa qualidade e 400 dpi para documentos de má qualidade.
4. Coloque o digitalizador na seção do documento que pretende reconhecer.



5. Pressione uma vez a tecla Start do digitalizador, em o movendo na direção selecionada.

A imagem é exibida progressivamente à medida que o digitalizador se move.

O movimento precisa ser lento e tão regular quanto possível, caso contrário a imagem varrida poderá ficar inutilizada para o OCR.

As instruções do fabricante fornecidas com o digitalizador lhe fornecem maiores informações sobre como usar o digitalizador.

6. Após ter varrido a imagem, pressione a barra de espaço para passar o controle para o Readiris.

Se a direção de varredura selecionada era da esquerda para a direita ou da direita para a esquerda, pressione a barra de espaço para rodar diretamente a imagem. Você também poderá rodar as imagens em seguida com o comando Rotate do menu View.

Preste atenção ao **desalinhamento**. Embora a análise de página e o reconhecimento efetuado pelo Readiris tolerem desvios, pode ficar difícil representar uma janela e fazer o reconhecimento ótico correto de uma página varrida quando o desalinhamento é demasiado significativo.

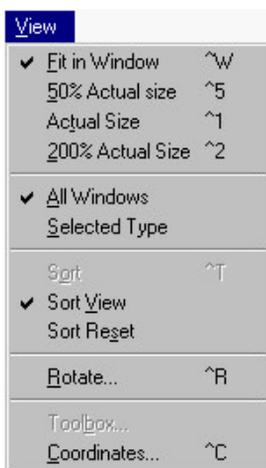
OLHANDO DE PERTO A IMAGEM VARRIDA

Exibir toda a página é útil para ver como os blocos de texto e os blocos gráficos se relacionam entre si. Essa visualização total da página se chama de Fit in Window.

Para controlar a qualidade da varredura do documento, você precisa olhar com maior detalhe. O Readiris fornece outros três tipos de visualização:

- Actual Size: aproxima o tamanho da página ao tamanho da folha de impressão
- 50% Actual Size: exibe a página a metade do tamanho atual

- ❑ 200% Actual Size: exibe a página duas vezes maior que o tamanho atual



Você também poderá alternar entre Actual Size e Fit in Window e vice-versa clicando o botão *direito* do mouse.

CONTROLANDO A QUALIDADE DA IMAGEM VARRIDA

O desempenho do reconhecimento ótico de caracteres (OCR) depende muito da qualidade do documento e da varredura. Para controlar essa qualidade, você precisa olhar para a imagem varrida com maior detalhe.

O Readiris produz melhores resultados em um documento de “boa” qualidade, por exemplo, documentos com caracteres bem impressos e razoavelmente bem separados.

Em um documento demasiado claro, os caracteres são, às vezes, cortados em duas partes: um "O" pode transformar-se em "()". Por outro lado, um documento demasiado escuro poderá conter letras com formas



demasiado grandes e as letras abertas poderão aparecer fechadas após a varredura: por exemplo, um "c" se torna um "o". Em um documento escuro, muitos caracteres podem ficar colados, impedindo o correto reconhecimento.

Pode melhorar a qualidade da imagem varrida, ajustando o brilho e o contraste do digitalizador.

Após uma nova varredura da imagem, você pode verificar as alterações na tela.

FERRAMENTAS DE REPRESENTAÇÃO EM JANELA

Quando o documento é exibido, o passo seguinte é indicar as zonas de interesse.

Existem várias formas de o fazer: a análise de página detecta e classifica automaticamente os diversos blocos. Você pode alterar o layout resultante com o botão Sort, definir manualmente as janelas e salvar estruturas de representação em janela específicas, chamadas de "layouts", em arquivos, para uso posterior.



Desenhar ou detectar diversas janelas é essencial toda vez que um texto está formatado em **colunas**. Para reconhecer corretamente tabelas, desenhe janelas à roda de cada coluna para organizar corretamente o resultado em células.

Eis uma descrição das ferramentas de representação em janelas do Readiris.

Criando uma janela retangular

1. Selecione a ferramenta do retângulo, clicando a ferramenta retangular na caixa de ferramentas Windows.



2. Selecione o tipo de janela: texto ou gráficos.
3. Mova o cursor para um canto da janela que pretende criar e clique.
4. Aumente a janela movendo o cursor para o canto oposto e clique. A janela é desenhada.

Não são permitidos lados menores que 1 mm!

Criando uma janela poligonal

1. Selecione a ferramenta do polígono, clicando a ferramenta poligonal na caixa de ferramentas Windows.
2. Mova o cursor para um canto da janela poligonal que pretende criar e clique.
3. Mova o cursor. Um dos lados do polígono é desenhado à medida que o mouse se move.
4. Clique para acabar de desenhando o primeiro lado.
5. Mova o cursor e clique onde pretende terminar um lado do polígono.
6. Clique o botão *direito* do mouse para fechar o polígono.

Você não pode cruzar um lado já desenhado!

Também pode criar polígonos mesclando retângulos. Toda vez que duas janelas do mesmo tipo se tocam, elas são mescladas para formar uma única janela.

Selecionando janelas

1. Selecione a ferramenta de seta na caixa de ferramentas Windows.
2. Clique a borda de uma janela.



Aparecem marcadores em cada canto e no meio dos lados da janela.

3. Para selecionar uma segunda janela sem desmarcar a primeira, pressione sem liberar a tecla Shift enquanto clica a segunda janela.
4. Para selecionar uma janela e todas as janelas nela incluídas, pressione sem liberar a tecla Ctrl enquanto clica a janela principal.
5. Para desmarcar janelas, clique o botão esquerdo do mouse fora das janelas.

Você pode ver o tipo e as **coordenadas** de uma janela pressionando a tecla Alt enquanto seleciona a janela.

Modificando uma janela

1. Selecione a janela. Ver acima.
2. Mova o cursor para um dos marcadores.
3. Clique e segure o botão esquerdo do mouse.
4. Mova o mouse.

A dimensão da janela se altera com os movimentos do mouse.

5. Libere o botão quando a dimensão da janela for a pretendida.

Movendo janelas

1. Selecione a(s) janela(s). Ver acima.
2. Mova o cursor para um dos lados de uma janela, em o colocando no meio de dois marcadores.
3. Clique e segure o botão do mouse.
4. Mova o mouse. A posição da janela muda consoante os movimentos do mouse.
5. Libere o botão quando a janela estiver na posição adequada.

Eliminando janelas

1. Selecione a(s) janela(s). Ver acima.

2. Selecione o comando Delete ou Clear no menu Edit.

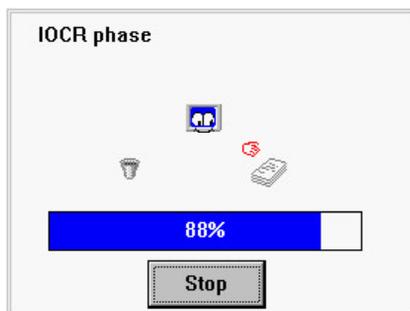
O comando Delete elimina a(s) janela(s) mas salva uma cópia interna que poderá ser colada posteriormente. O comando Clear elimina definitivamente a(s) janela(s).

RECONHECIMENTO DE CARACTERES

Clique o botão OCR na barra de botões para reconhecer caracteres nas zonas indicadas, mesmo que pretenda somente salvar gráficos.



O progresso é exibida na tela e pode ser interrompido em qualquer momento clicando o botão Stop.

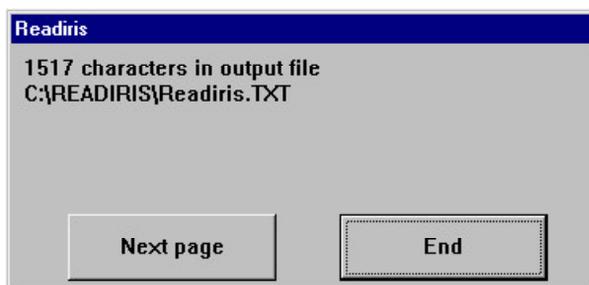


Uma vez que o Readiris suporta **multitarefa**, você também pode executar o reconhecimento em segundo plano. Minimize a janela Readiris e efetue outra tarefa no seu PC; a janela de aprendizagem interativa é exibida quando o programa está pronto.

Durante essa fase interativa, o texto reconhecido é exibido progressivamente. Quando o sistema não tem certeza quanto a uma solução, a imagem da palavra é exibida e o caractere duvidoso fica destacado. O sistema pára aguardando sua resposta. Ver abaixo.



Após a fase de validação, o texto reconhecido é salvo no arquivo de saída ou na área de transferência e o sistema pergunta se você deseja ler outra página.



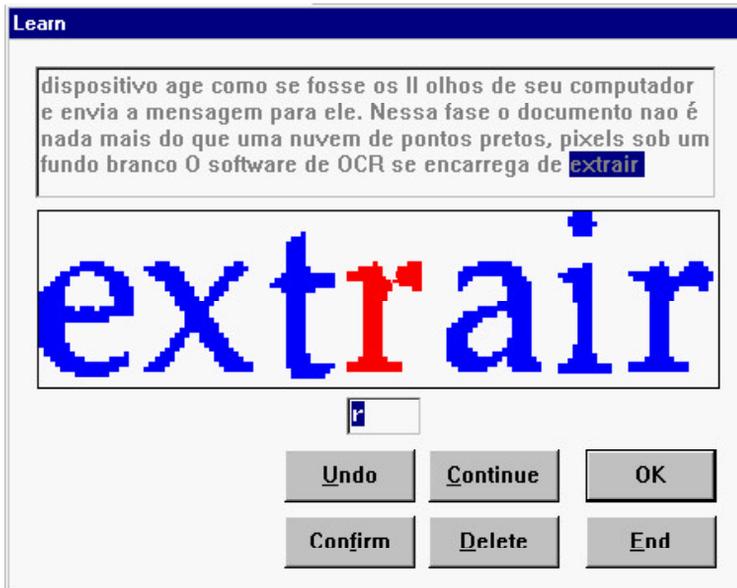
No caso de documentos com várias páginas, o botão Next Page coloca automaticamente o dicionário de fontes e o arquivo de saída em modo Append. Nos digitalizadores planos, a página seguinte é varrida e nos digitalizadores portáteis, você pode varrer de novo.

Facilmente você verá que a fase de aprendizagem precisa menos interação nas páginas seguintes, quando o fonte é semelhante.

APRENDENDO SOBRE FONTES

A fase final do reconhecimento é a **fase de aprendizagem**. (O nível de interação depende da opção Strong Checking no Control Panel.)

Se o Readiris não tem certeza quanto a um resultado do reconhecimento, o caractere duvidoso é exibido com a palavra à qual pertence. Esse caractere - ou cadeia de caracteres, se não estão separados - é destacado, sendo exibida a solução proposta. Os caracteres não reconhecidos são assinalados como "~".



Se necessário, escreva os caracteres válidos e selecione um dos seguintes botões.

OK

Você pode aceitar a solução proposta ou a corrigir. O resultado é armazenado no dicionário de fontes como certo e definitivo: sua intervenção não será necessária em futuros reconhecimentos.

Alternativa: pressione Enter.



Continue

Você pode aceitar a solução proposta ou a corrigir. A diferença em relação ao OK é que o Readiris salva o resultado como informação “incerta” no dicionário de fontes.

Em reconhecimentos futuros, sua intervenção ainda será necessária.

Use esse comando para caracteres danificados, que poderiam ser confundidos com outros caracteres se fossem aceitos.

Confirm

O número '1' e a letra "l" têm uma forma idêntica em muitas faces de tipos. Nesse caso, o Readiris faz uma análise de contexto para os diferenciar. Quando o contexto não é suficiente, o usuário terá que decidir.

Use esse botão quando as formas para o número '1' e para a letra "l" são diferentes, caso contrário, use Continue.

Undo

Permite retornar para corrigir erros. O Readiris mantém um registro das últimas nove operações!

Delete

Elimina caracteres da saída. Use esse botão para evitar que as manchas do documento sejam reconhecidas como pontos, vírgulas, etc.

End

Aborta o processo de aprendizagem interativa. A fase de decisão será completada automaticamente. Você aceita todas as sugestões do Readiris.

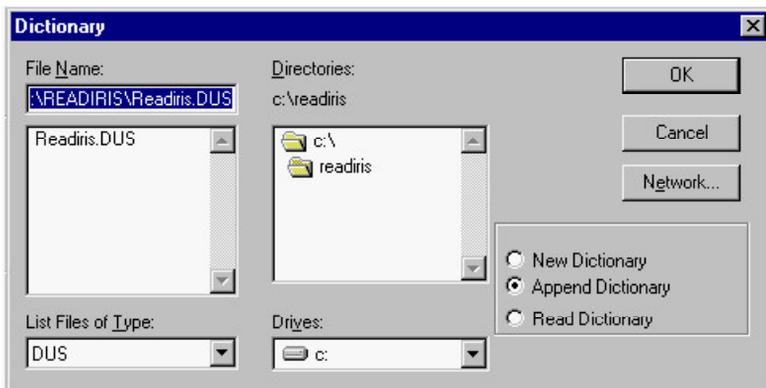
USANDO CORRETAMENTE OS DICIONÁRIOS DE FONTES

Se você executar o Readiris em vários documentos do mesmo tipo, poderá não desejar repetir sempre o mesmo processo de aprendizagem. Aí, você pode salvar os caracteres aceitos em um dicionário de fontes e usar esse conhecimento adquirido quando efetuar o reconhecimento ótico de documentos semelhantes.

Para criar ou ativar um dicionário de fontes:

1. Selecione Control Panel no menu Process.
2. Clique Dico.

São listados todos os dicionários de fontes existentes no diretório atual.



3. Escreva um nome para criar um novo dicionário de fontes ou selecione um já existente.
4. Clique um dos modos disponíveis: New para criar um dicionário de fontes, Append para completar um dicionário existente e Read para usar um dicionário de fontes sem salvar informações novas.
5. Clique OK.



6. O arquivo do dicionário ativo é exibido junto do botão Dico no Control Panel. O modo selecionado é indicado entre parênteses.

AJUDA

Esta seção termina a panorâmica geral sobre o funcionamento do Readiris. O Readiris contém um sistema de **ajuda on-line** que documenta exaustivamente cada aspectos do software. A ajuda on-line também poderá conter algumas informações de última hora.

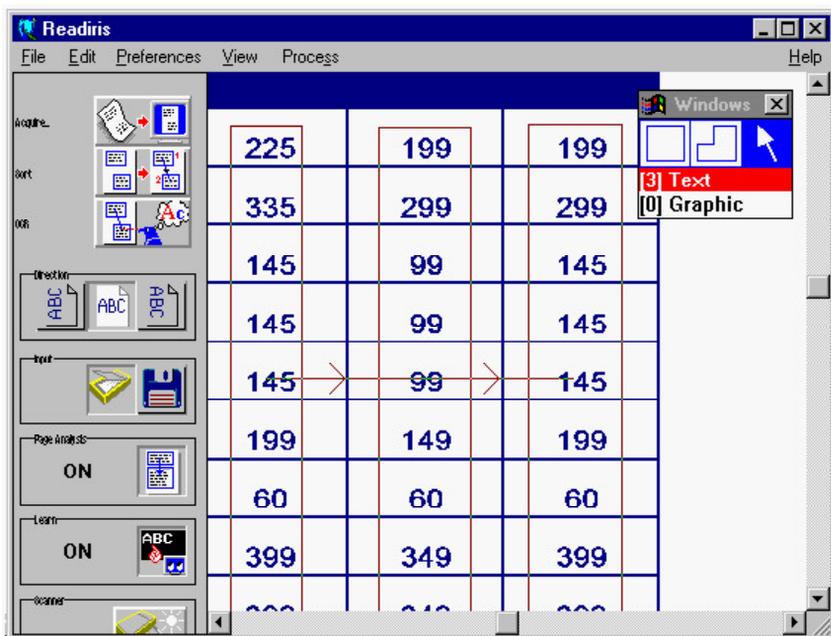
Selecione o comando Contents no menu Help para acessar essas informações. Atalho: pressione F1. A opção About Readiris exibe a tela inicial com seus dados de registro.

Entretanto, o Readiris contém algumas **funções avançadas** que discutiremos mais detalhadamente. Vamos explicar como reconhecer tabelas, como executar o reconhecimento ótico de caracteres (OCR) automático e como usar a capacidade de conexão.

RECONHECENDO TABELAS

Você pode usar o Readiris para importar tabelas diretamente para sua planilha.

É obrigatório desenhar uma janela à roda de cada coluna para organizar o resultado em células correspondentes, caso contrário não poderá importar as tabelas corretamente. A figura abaixo é um bom exemplo.



Selecione um dos dois formatos de texto seguintes: ASCII do Excel ou ASCII da área de transferência do Excel. Embora ambos forneçam um filtro de tabelas genérico, o formato ASCII do Excel cria um arquivo de tabelas que poderá ser importado diretamente para a planilha, enquanto que o formato da área de transferência do Excel envia a tabela reconhecida para a área de transferência do Windows. Use o comando Paste da planilha para inserir a tabela.

A detecção de parágrafos e a edentação são desativados toda vez que selecionar um formato de tabela.



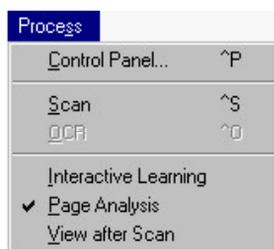
O Readiris pode limitar o reconhecimento a um conjunto de caracteres numéricos quando as tabelas de figuras sem texto são reconhecidas opticamente (OCR). Selecione o **modo numérico** na lista de cortina Language da barra de botões.



RECONHECIMENTO AUTOMÁTICO

Você pode reconhecer automaticamente suas páginas. O modo automático como o reconhecimento é executado depende do tipo de digitalizador.

Os **digitalizadores planos** permitem o reconhecimento automático de várias páginas. Desative as opções Interactive Learning e View after Scan e ative a opção Page Analysis. (Você também pode usar um layout de representações em janela para indicar as zonas de interesse.) Escreva o número de páginas a reconhecer no Control Panel.



Coloque um conjunto de documentos no alimentador de documentos do digitalizador, clique o botão Start para iniciar a varredura e deixe seu PC fazer o restante. Os documentos são varridos, zonados e reconhecidos sem você precisar intervir! Se o alimentador de documentos ficar vazio antes do reconhecimento do número de páginas especificado, é exibida uma mensagem de erro.

Com os **digitalizadores portáteis** é diferente. Uma vez que a varredura implica sempre algum esforço físico do usuário, o reconhecimento nunca é totalmente automático. Apesar disso, você poderá automatizar as fases de representação em janelas e de reconhecimento.

Dado o reconhecimento automático implicar a desativação da fase de aprendizagem, primeiro você deve testar o Readiris em algumas páginas, caso contrário, o resultado poderá não mostrar um nível de reconhecimento ótimo.

USANDO A CAPACIDADE “CONNECT” (CONEXÃO)

A capacidade de conexão oferece uma ligação direta OCR entre o digitalizador e o aplicativo do Windows. Pode até reconhecer textos oticamente a partir do seu aplicativo do Windows favorito!

Após instalar a função Connect, é exibido um novo comando nos menus do aplicativo do Windows. Clique esse comando para executar o Readiris. Pode varrer e reconhecer seu documento e o resultado é introduzido diretamente no aplicativo do Windows, na posição atual do cursor.

WordPerfect, Word e Excel

O Readiris não pode ser executado em segundo plano.

Clique o novo comando Scan Text no menu do aplicativo para executar o Readiris no aplicativo do Windows. No WordPerfect 5.x, este comando é exibido no menu Macro, no WordPerfect 6.0 no comando Macro do menu Tools e no Word e Excel no menu File.

O interface de conexão poderá fazer algumas perguntas antes da execução do Readiris. Proceda de forma normal e saia do interface de conexão quando terminar. Os resultados do reconhecimento são introduzidos na posição atual do cursor.

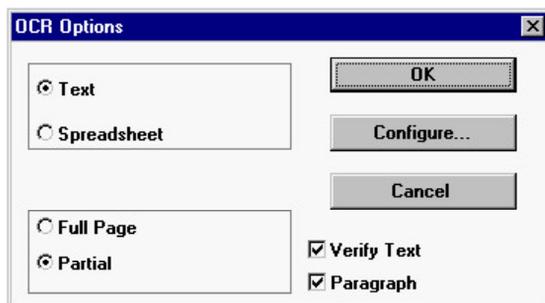


Outros aplicativos do Windows

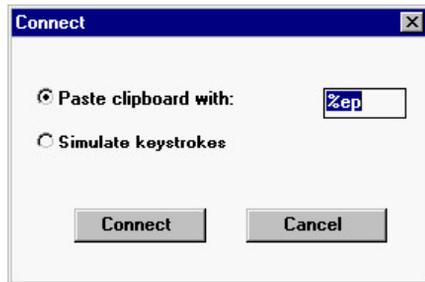
Clique duas vezes o ícone Connect para executar o servidor de tarefas Connect em segundo plano *antes* de executar o aplicativo do Windows. Clique o novo comando OCR no menu File do seu aplicativo. É exibida uma caixa de diálogo.

Indique se vai ler um texto (Text) ou uma tabela de números (Spreadsheet). Selecione Full Page para ativar a análise do total da página ou Partial para representar em janela manualmente o documento varrido.

Ative a opção Verify Text para incluir a fase de aprendizagem interativa. A validação do usuário é desativada por defeito.



O botão Configure permite acessar alguns parâmetros avançados. Selecione Simulate keystrokes para introduzir o texto reconhecido como tendo sido digitado a partir do teclado ou selecione Paste Clipboard para colar o texto de forma invisível a partir da área de transferência.



Verifique se a tecla de atalho correta está especificada para o comando de colagem do aplicativo do Windows.

Algumas teclas são representadas por um caractere especial.

Tecla	Representação
Shift	+
Ctrl	^
Alt	%
Insert	{Insert}

Exemplos: se a tecla de atalho para o comando Paste é Ctrl+V, você deve escrever **^V**, se a tecla de atalho é Shift+Insert deve escrever **+{Insert}**, etc.



4. REFERÊNCIA A COMANDOS

Este capítulo é um breve guia de referência de todos os menus do Readiris. Se você já leu os capítulos anteriores, estará já familiarizado com esses comandos.

MENU FILE

Permite salvar e recuperar definições e arquivos de representação em janela. Permite abrir uma imagem previamente varrida.

Open

Abre um arquivo de imagens de formato TIFF. Ativo apenas no modo Preview.

Get Info

Exibe a configuração ativa do Readiris.

Select Source

Estando o Readiris em conformidade com a norma de digitalizadores Twain, este comando permite selecionar um digitalizador Twain e definir os parâmetros de varredura.

Acquire

Inicia a varredura. Idêntico ao comando Scan do menu Process e ao botão Scan da barra de botões e do Control Panel.



Save Layout

Salva a estrutura de representação em janela tal como definida em um documento durante sua prévia. Os layouts são salvados em um arquivo com a extensão predefinida *.wdw. Ativo apenas no modo Preview.

Get Layout

Recupera as estruturas das janelas previamente armazenadas em um arquivo com a extensão predefinida .wdw. Ativo apenas no modo Preview.

Save Settings

Salva a configuração atual do Readiris em um arquivo com a extensão predefinida .wdw.

Get Settings

Recupera a configuração do Readiris previamente armazenada em um arquivo de configuração.

Set Default Settings

Armazena a configuração atual como predefinição para inicialização do Readiris.

Exit

Sai da sessão do Readiris.

MENU EDIT

Os comandos desse menu só estão ativos durante a prévia do documento: atuam em janelas definidas quando o documento é exibido na tela.

Para selecionar uma janela, clique um lado da janela. Se você quiser selecionar simultaneamente mais janelas, pressione sem liberar a tecla Shift quando clicar os lados das janelas.

Undo

Cancela o comando anterior.

Delete

Elimina as janelas selecionadas, salvando as mesmas em um buffer interno.

Copy

Copia as janelas selecionadas para um buffer interno.

Paste

Insera o conteúdo do buffer interno. As janelas inseridas aparecem no local onde foram copiadas ou eliminadas. Para mover as janelas, clique o lado de uma dessas janelas e arraste o mouse: as janelas são movidas consoante os movimentos do mouse.

Clear

Elimina definitivamente todas as janelas selecionadas.

Select All

Seleciona simultaneamente todas as janelas.



MENU PREFERENCES

Especifica as definições básicas do Readiris: qual o formato de texto e gráfico usado e o idioma do documento a ser lido. Ativa a barra de botões.

Scanner

Exibe o tipo do digitalizador e as respectivas definições.

Text Format

Seleciona o formato do arquivo para os caracteres reconhecidos.

Ativa a detecção de parágrafos e a codificação das margens. Quando você fizer o reconhecimento de tabelas, não esqueça de selecionar o formato Excel.

Graphics Format

Seleciona um formato gráfico para as zonas gráficas. Verifique quais os formatos de arquivo aceitos pelo aplicativo de gráficos.

Language

Seleciona o idioma do documento. O Readiris está equipado com vários bancos de dados lingüísticos. Você também pode limitar o reconhecimento de caracteres a caracteres numéricos, por exemplo, para reconhecer tabelas numéricas sem texto com o máximo de precisão e velocidade.

Toolbar

Ativa a barra de botões. A barra de botões é exibida no lado esquerdo da janela do aplicativo. Recomendamos que use a barra de botões porque esses botões permitem o acesso rápido às funções de menu mais usadas.

MENU VIEW

Os comandos desse menu se encontram disponíveis somente durante a prévia do documento.

Fit in Window

Ajusta o nível do zoom para a imagem do documento caber na tela.

50% Actual Size

Ajusta o nível de zoom para 50%.

Actual Size

Faz um zoom até à dimensão real do documento: um pixel da tela iguala um pixel de imagem. Pode alternar entre Actual Size e Fit in Window e vice-versa, clicando o botão *direito* do mouse.

200% Actual Size

Ajusta o nível de zoom para 200%.

All Windows

Exibe todas as janelas definidas no layout do documento.

Selected Type

Exibe somente as janelas do tipo selecionado.

Sort

Classifica as janelas de texto. A ordem da classificação determina a ordem de exibição das diferentes janelas de texto. Esta opção é bastante útil para documentos com várias colunas. Clique em uma borda da janela para definir a ordem de classificação dos vários blocos de texto.



Sort View

Exibe a seqüência lógica das janelas como definida no comando Sort. As setas ligam as diferentes janelas de texto.

Sort Reset

Restaura a ordem de classificação das janelas texto. Alternativa: use o botão Sort da barra de botões.

Rotate

Roda a imagem varrida. Selecione a rotação da imagem no sentido dos ponteiros do relógio ou no sentido contrário. Quando fizer a varredura, selecione a direção adequada.

Toolbox

Visão da caixa de ferramentas Windows que permite selecionar o tipo de janela (texto ou gráficos) e a ferramenta da janela (retângulo, polígono ou seta).

Coordinates

Exibe a janela de coordenadas. As coordenadas do cursor são exibidas em pixels de documento.

MENU PROCESS

Contém os comandos relacionados com o atual processo OCR: varredura, reconhecimento ótico e todos os parâmetros relativos a essas operações.

Control Panel

Exibe o Control Panel.

Scan

Inicia a varredura de um documento ou abre uma imagem previamente varrida, dependendo da fonte de entrada selecionada. Equivalente ao botão Scan da barra de botões.

OCR

Inicia o reconhecimento de caracteres e salva os gráficos. Equivalente ao botão OCR da barra de botões.

Interactive Learning

Ativa a aprendizagem interativa. Determina se você deve verificar os caracteres duvidosos ou aceitar as soluções do sistema. (A opção Strong Checking do Control Panel determina o nível de verificação pelo usuário.)

Você também pode acessar o modo de aprendizagem através da barra de botões. Desative esta opção para executar o reconhecimento ótico de caracteres (OCR) automático.

Page Analysis

Ativa a análise de página. Determina se a varredura é seguida da decomposição automática da página.

Você também pode ativar e desativar a Page Analysis a partir da barra de botões.

Ative esta opção para executar o reconhecimento ótico de caracteres (OCR) automático.

View after Scan

Ativa o modo Preview: o documento varrido é exibido na tela. Você pode definir as diferentes partes do texto a reconhecer e os gráficos a extrair pelo Readiris. Se tiver um digitalizador portátil, não desative o modo Preview.



CONTROL PANEL

As especificações de Control Panel estão descritas detalhadamente no capítulo 3 “Aprendendo mais sobre o Readiris”.

Pitch

Seleciona a densidade horizontal: fixa ou proporcional.

Size

Seleciona o tamanho dos caracteres: pequenos ou normais. Um caractere é considerado “pequeno” se seu tamanho em ponto for menor que 10 ou se o software tiver problema para separar corretamente duas linhas adjacentes.

Font

Indica a qualidade de impressão do texto: você lê fontes “normais” ou textos impressos em matriz de pontos.

Pages(s) to Read

Número de páginas a ler automaticamente. Esta opção não está disponível nos digitalizadores portáteis.

Page Analysis

Ativa a análise de página. Determina se a varredura é seguida da decomposição automática da página.

Você também pode ativar a Page Analysis a partir da barra de botões.

Ative esta opção para executar o reconhecimento ótico de caracteres (OCR) automático.

View after Scan

Ativa o modo Preview: o documento varrido é exibido na tela. Equivalente a View after Scan no menu Process.

Interactive Learning

Ativa a aprendizagem interativa. Determina se você deve verificar os caracteres duvidosos ou aceitar as soluções do sistema.

Você também pode acessar o modo de aprendizagem na barra de botões. Desative esta opção para executar o reconhecimento ótico (OCR) automático. Também pode influenciar o nível de interação com a opção seguinte.

Strong Checking

Quando esta opção está ativada, o nível de “rejeição” é elevado: o Readiris confrontará o usuário com as menores dúvidas, ao invés de tomar algumas decisões autonomamente. Quando o Readiris reconhece um texto sem problema, você pode verificar somente as formas que constituem uma dúvida real.

Input

Selecione a fonte de entrada. Você executa um reconhecimento ótico de caracteres (OCR) a partir do digitalizador ou a partir de um arquivo de imagens previamente varridas.

Output

Define o nome do arquivo de saída. Se esse arquivo já existe, você precisa selecionar o modo Replace ou Append. A extensão do arquivo de saída é acrescentada automaticamente, dependendo da base do formato de saída selecionado.



Layout

Seleciona o arquivo de layout que contém as estruturas das janelas adaptadas a um tipo de documento. A extensão predefinida é *.wdw.

Dico

Selecione o dicionário de fontes e seu modo: "New" para criar um novo dicionário de fontes, "Append" para completar um dicionário existente e "Read" para usar e não escrever no dicionários de fontes. Por predefinição, a extensão do arquivo é *.dus.

Scan

Inicia a varredura. Equivalente ao botão Scan da barra de botões.

Close

Fecha o Control Panel.

MENU HELP

Este menu contém o sistema de ajuda on-line e a janela About.

Contents

Permite acessar o sistema de ajuda on-line que documenta exaustivamente cada aspecto do software. A ajuda on-line também poderá conter informações de última hora.

Tecla de atalho: pressione F1.

Você pode consultar a informação na tela e impressa! O guia do Windows lhe diz tudo o que você precisa fazer.

About Readiris

Exibe a tela inicial com os dados de registro.

5. MAIS SOBRE OCR

O objetivo do OCR é introduzir automaticamente documentos impressos e dactilografados de forma eficaz e econômica. Embora a primeira pesquisa e desenvolvimento do Reconhecimento Ótico de Caracteres (OCR) tenha começado há mais de 20 anos, esta tecnologia continua desconhecida para a maioria das pessoas que a poderiam, e deveriam, usar em seus aplicativos de entrada de documentos.

Você pode instalar esta poderosa ferramenta em seu escritório para evitar a tarefa fastidiosa de escrever novamente o texto impresso.

Os sistemas OCR funcionam bem para vários tipos de documentos, mas não para todos... As páginas seguintes descrevem brevemente os diferentes passos para fazer o reconhecimento ótico (OCR) de um documento. Também é fornecida uma panorâmica geral das especificações do Readiris.

PROCESSO DE RECONHECIMENTO ÓTICO (OCR)

O documento é lido pelo digitalizador. O dispositivo eletrônico atua como o “olho” de seu computador em lhe enviando a imagem. Nesta fase, a imagem do documento não é mais que uma nuvem de pontos pretos sem sentido, chamados de “pixels”, em um fundo branco. O software do OCR precisa extrair informações de textos desse pixels: precisa reconhecer as formas e atribuir um símbolo.

Existem dois tipos de sistemas OCR: sistemas “automáticos” e de “aprendizagem”. O primeiro tipo pode reconhecer texto automaticamente a partir de várias fontes... mas nem todas. Com o segundo tipo, você precisa ensinar novos caracteres ao sistema toda vez que tem um novo



conjunto de documentos. Este método necessita mais interação, mas pode tratar quase todas as fontes.

Ambos os sistemas OCR usam processos equivalentes para reconhecer texto. Os passos principais são:

- segmentação de linhas
- segmentação de palavras e caracteres
- reconhecimento de caracteres
- saída de texto

A **segmentação de linhas** consiste em dividir uma página de texto em diferentes linhas. Este passo também analisa o desalinhamento e o espaço entre linhas. Os documentos mais difíceis são os que têm um pequeno espaçamento entre linhas e um desalinhamento significativo.

A **segmentação de palavras e caracteres** isola as palavras umas das outras e depois separa as diferentes letras de cada palavra. Se todos os caracteres tiverem a mesma largura (fontes fixos), fica fácil executar este passo. Isso acontece, por exemplo, em documentos dactilografados em uma máquina de escrever.

O problema é mais interessante quando a largura das letras varia conforme sua forma atual (fontes proporcionais). Os casos mais difíceis são os documentos onde os caracteres proporcionais raramente se tocam. Os caracteres de matriz de pontos são compostos por pequenos pontos e necessitam de técnicas de segmentação muito específicas.

O **reconhecimento de caracteres** extrai as características de cada forma isolada dando um nome a cada uma.

A **saída de texto** é a conversão do texto reconhecido em um arquivo que pode ser importado para o processador de texto, planilha, software de editoração eletrônica, etc.

ÂMBITO DO READIRIS

O Readiris é um programa de OCR bem fácil de usar e de grande desempenho. Suas funções foram selecionadas a partir das técnicas mais avançadas IORC, controladas pela IRIS. Este produto tem as seguintes especificações:

- ❑ Usa as **técnicas de aprendizagem automática** para ser um sistema aberto. O Readiris é um sistema de aprendizagem que parece ser automático: o software encontra a solução certa para a maior parte do documento.
- ❑ O sistema usa exaustivamente o **contexto lingüístico** para encontrar as soluções, mas também pode funcionar sem qualquer contexto, por exemplo, em tabelas numéricas.
- ❑ Guia o usuário de forma simples através de um **processo composto por cinco passos**: varredura, prévia, representações em janelas, reconhecimento e validação dos resultados duvidosos.

Tal como outros produtos OCR, o Readiris não suporta *todos* os tipos de documentos, mas o que ele faz, faz bem feito. Este sistema confiável e econômico se destina a ler documentos originais e/ou cópias de boa qualidade.

As capacidades de aprendizagem são muito poderosas e permitem obter os melhores resultados se as seguintes condições forem respeitadas.

Número de caracteres e contexto lingüístico

O Readiris necessita um número razoável de caracteres e de contexto lingüístico para cada símbolo representado, a fim de executar o processo de aprendizagem automática de tal modo que o reconhecimento seja quase todo automático.



Se o contexto lingüístico é insuficiente ou não existe, o sistema funciona ainda assim embora com maior interação do usuário.

Isolamento de caracteres

O Readiris executa um algoritmo de segmentação limitado para reduzir o tempo de processamento e os requisitos de memória.

O primeiro passo para o isolamento de caracteres pode ser executado em caracteres um pouco separados. Em um segundo passo, o Readiris usa um algoritmo muito poderoso para decompor grupos de caracteres espaçados e ligaduras aleatórias.

Desse modo, o Readiris precisa ter isolado um mínimo de modelos elementares (os caracteres isolados) para executar uma segmentação correta dos caracteres que se tocam.

Consequentemente, o reconhecimento de qualquer fonte é possível contanto que o sistema isole um número suficiente de caracteres elementares.

Gama de tamanho de caracteres

A gama de tamanhos é de 8 a 20 pontos, o que abrange a maioria dos tamanhos de fontes usados para o corpo do texto. Entretanto, se você selecionar uma resolução de 400 dpi, mesmo as fontes menores serão legíveis.

Não misture um título de algumas palavras com um tipo de letra maior com o texto a reconhecer, que é uma parte muito maior. Na maioria dos casos, o operador de DTP escreverá seus próprios títulos.

Documentos excluídos

O Readiris não lê textos demasiado densos, ou seja, textos onde praticamente todas as letras de cada palavra se tocam. Esse é o motivo para a exclusão de documentos de má qualidade, como as fotocópias muito escuras.



PRODUCT REGISTRATION

Don't forget to send us your registration card! Should your **registration card** have gotten lost, please fill out this card and send or fax it to I.R.I.S.

Your registration will allow us to keep you informed of related I.R.I.S. **products** and new **developments**. Registering also entitles you to **product support** and **special offers**.

Product / version

Serial number

Company

Name

Title

Street / number

ZIP code / city

State / country

Telephone

Fax

Where and when did you buy?

Date

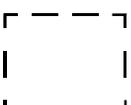
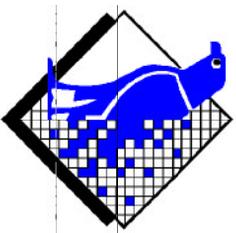
Image Recognition Integrated Systems

Product registration

Rue du Bosquet 10, 1348 Louvain-la-Neuve (Belgium)

Tel: 32-10-45 13 64 - Fax: 32-10-45 34 43





I.R.I.S.

Product registration
Rue du Bosquet 10,
1348 Louvain-la-Neuve
Belgium