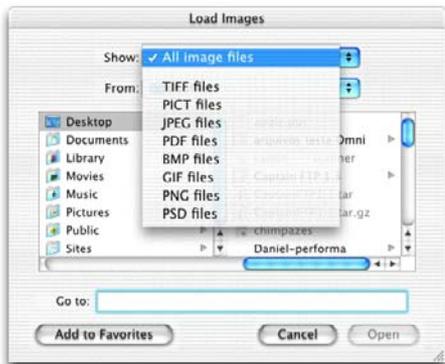
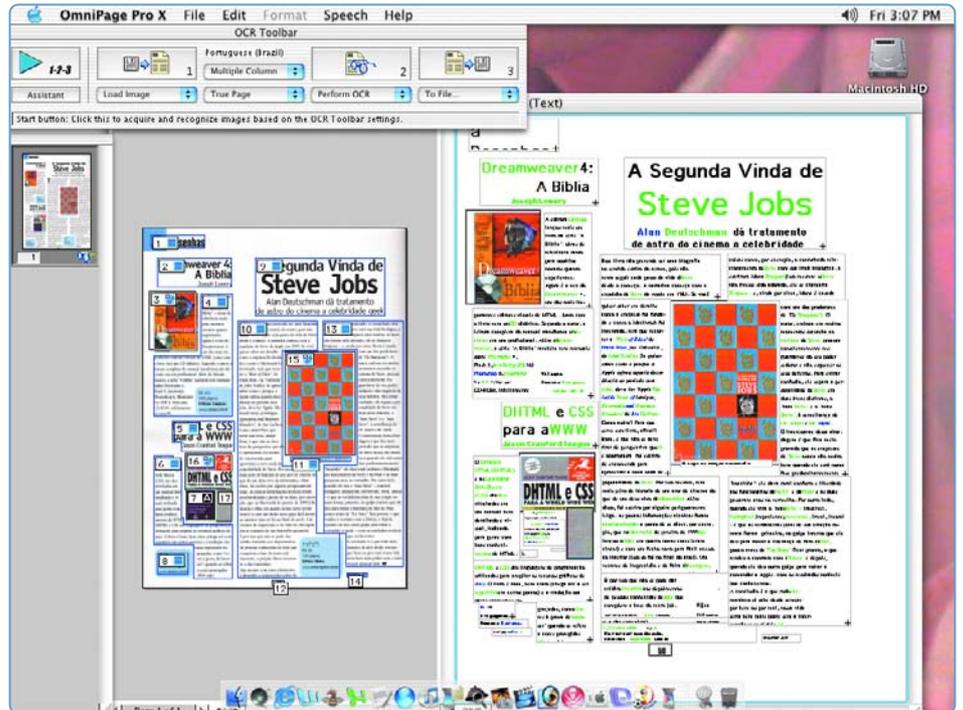




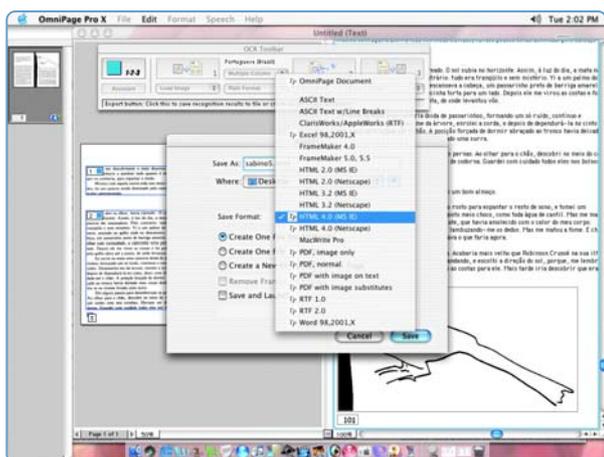
Quem nunca teve que redigitar páginas e páginas de texto para um trabalho de escola, arquivo de notícias ou qualquer outro motivo? Esse trabalho pode ser poupado com um bom programa de reconhecimento de caracteres (conhecido pela sigla OCR, de *optical character recognition*). Mas quem já usou um software desses sabe que eles não costumam ser confiáveis. Muitas vezes é mais fácil digitar tudo do que perder um tempão arrumando o texto porque nem a metade foi “entendida” pelo computador. Por essas e outras razões, ainda é baixa a aceitação deste tipo de aplicativo, principalmente por pequenas empresas e usuários domésticos. A nova versão para o Mac OS X do OmniPage Pro, da ScanSoft, dá uma renovada nos turvos ares dessa área de programas. Além da vantagem natural de funcionar a contento no OS X (ele também roda no sistema 9), o OmniPage melhorou muito se comparando com a versão 8. Não só em relação ao reconhecimento da escrita, mas em outros aspectos, como facilidade de uso e suporte a mais formatos de arquivos. Entretanto, continua apresentando alguns problemas antigos. A intuitividade do OmniPage merece um aparte: o mais lesado dos pokaprátikas consegue em dez

OmniPage Pro X

Transformar textos em pixels ficou mais fácil e melhor, mas ainda distante da perfeição



Grande compatibilidade com formatos de arquivos

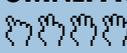


minutos dominar com esmero o programa. As funções mais importantes estão na barra de ferramentas principal. Há até um botão que faz automaticamente todos os passos básicos de reconhecimento, que consistem em distinguir o que é texto do que é imagem e depois transformá-lo em caracteres. A janela principal é dividida em duas partes: na esquerda aparece a imagem original e no outro lado, o texto reconhecido. Mais simples, impossível.

Existe a possibilidade de treinar a interpretação da escrita, que pode ser gravada em vários arquivos. Ideal quando se quer reconhecer uma grande quantidade de texto com o mesmo tipo de formato e fonte, como num livro, por exemplo. Essa função evita, na maioria das vezes, que letras sejam “canibalizadas” ou que todo ‘n’ se transforme em um acento – o mais perverso problema dos programas de OCR. Mesmo assim, o programa continua tendo dificuldades com

Devidamente atualizado para o Mac OS X

acentos, e ainda depois de um bom treinamento. Talvez a explicação mais plausível seja que, como o inglês não usa acentos, os programadores anglo-saxões acabam tendo menos cuidado nesse aspecto. Não podemos deixar de lembrar de uma ferramenta importante incluída agora: a função de distinguir tabelas de texto, que nas versões anteriores saíam sem ordem de tabulação.

OMNIPAGE PRO X

 ScanSoft: www.scansoft.com
 11-5572-0247
 Preço: R\$3.018 (upgrade R\$555,97)

-  **Pró:** Bom reconhecimento de escrita; interface fácil de usar; lê em português
-  **Contra:** Problemas com quebras entre parágrafos; dificuldade em reconhecer automaticamente o que é texto ou imagem em páginas mais complexas; muito caro

EM MANTEN MANTEN

**Gamisetas da Macmania
e do Macintóshico
preta ou branca**



R\$ 20
Mais o frete

Ligue: 11-3341-5505

Mantenha a formatação original das tabelas

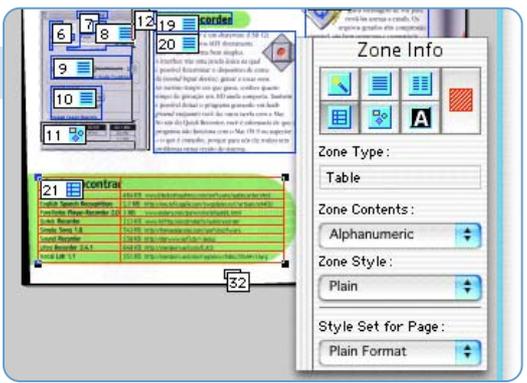
Jogo de cintura

As imagens que serão utilizadas podem ser carregadas direto do scanner (aceita-se qualquer tipo de scanner TWAIN) ou abertas de arquivos de imagens salvos no disco. O OmniPage lê com facilidade diversos arquivos padrão, como TIFF, PICT, JPEG, GIF, PSD e até PDF, entre outros. O programa também salva o texto em diversos formatos, como AppleWorks, HTML, PDF, Word e Excel. Em alguns, mantém até a disposição original dos textos e das imagens na página, jogo de cintura que a versão antiga não tinha.

A lista de dicionários é bem completa, incluindo o português do Brasil. Até línguas como Catalão ou Papiamento podem ser lidas, usando-se as línguas padrão como referência. No entanto, o léxico para a nossa língua é fraquinho se comparado ao dos outros dicionários para Mac disponíveis no mercado. O pulo do gato é que é possível incluir palavras novas e importar arquivos de palavras de outros programas, como o Excalibur (comentado na *Macmania 94*).

Reconhecendo na boa

Testamos o OmniPage X com o scanner Canon N670U, que tem plug-in para o OS X. Para o teste, usamos uma página de livro (uma coluna apenas e com fontes grandes, corpo 14) e três páginas diversas da Macmania (três e quatro colunas, duas fontes padrão em corpo 10). Todas as páginas foram escaneadas a 300 ppi (o mínimo aceitável pelo programa é de 200 ppi), algumas em preto e branco e outras em cores. Os arquivos foram salvos nos formatos TIFF, PICT e JPEG. Para abrir ou para escanear direto, não houve nenhum problema. O OmniPage “leu” todos os arquivos com facilidade. O reconhecimento do texto do livro foi excelente. Das cinco páginas, com uma média de 1500 toques cada, apenas um caractere “i” não



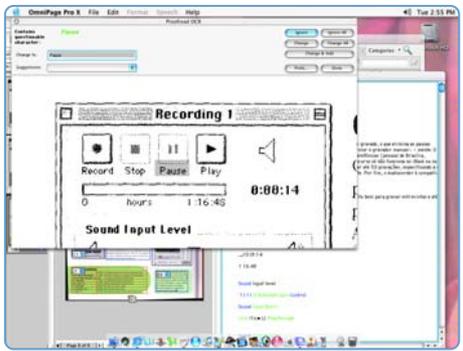
foi reconhecido. Por ser um texto relativamente simples, o corretor ortográfico não entendeu em média cinco palavras por página. Um desempenho expressivo.

Já nas páginas da Macmania, a coisa complicou um pouco. Por ter muitos termos técnicos, várias colunas e imagens misturadas, o programa se embananou para distinguir o que era texto, imagem ou tabela. Daí o maior pecado do OmniPage: muito do que era texto foi indicado como imagem e vice-versa, ou até colocado fora de ordem. Neste tipo de página mais complexa, o melhor é indicar no punho o que é cada um, mas aí o programa perde produtividade, já que, dependendo do número de páginas, você gasta boa parte do tempo arrumando isso. O que compensa é que as ferramentas para esse serviço são bem flexíveis e eficientes.

Quanto ao reconhecimento do texto em si, o programa teve uma média de acerto superior a 95%, com o corretor apanhando para desvendar as palavras, mas nada comprometedor. Há a opção de compor o texto final (aquele que fica no lado direito) na mesma posição do scan original. Neste tipo de caso, é uma função essencial. No entanto, ele “quebra” o texto no fim de cada coluna, não entendendo que uma é continuação da outra. É preciso fazer uma limpeza “na mão” para o texto ficar na ordem em que foi escrito. Outro argumento que prova que o programa ainda não chegou à perfeição. Enfim, o desempenho final do OmniPage vai depender muito da disposição do texto na página e da qualidade da imagem.

Trocando em miúdos, o OmniPage Pro X significa um bom avanço nos programas de OCR. Não é o ideal, já que exige bastante intervenção do usuário. Além de o resultado depender muito do formato da página. É uma compra ideal para quem quer digitalizar livros inteiros ou textos antigos datilografados, visto que não é lá muito barato. Para quem tem a versão 8 do programa (que vem de graça em muitos scanners), o upgrade é recomendável. **M**

Reconhecer o texto dentro da figura ou não?



DANIEL RONCAGLIA

Demorou uns dois anos para se formar na escola de datilografia.