



# O.C.R., minha datilógrafa



Valter Harasaki

Não, OCR não são as iniciais de Odete Calipígia Ramos, minha secretária. OCR quer dizer – mais uma sigla! – Optical Character Recognition (reconhecedor ótico de letras). Trata-se de uma categoria de softwares desenvolvidos para auxiliar as árduas tarefas de digitação, já que, comprovadamente, a maioria dos usuários de Mac são catadores de milho.

Com um scanner (de mão ou de mesa) e um texto, estes programas são capazes de reconhecer as letras do documento original e transformá-las em informações digitais que você pode utilizar ou modificar em seu microcomputador.

Porém, o fato de a maioria ter sido desenvolvida para a língua inglesa, alguns deixam a desejar, já que são problemáticos para reconhecer acentos e cedilhas (um “ç” pode ser confundido com “g”, por exemplo). Felizmente, nos programas mais sofisticados, é possível acrescentar novas letras para serem reconhecidas.

Mas será que vale a pena investir em um OCR? A resposta é: depende dos originais disponíveis. Isto porque, apesar do desenvolvimento notável que estes programas tiveram nos últimos anos, as variáveis que podem existir no documento original são quase infinitas. O tipo de letra do documento, o tamanho, o papel utilizado, se foi enviado por fax etc, influenciam na maior ou menor eficiência de reconhecimento.

Cada caso é um caso, por isso é preciso estar atento ao percentual de acertos que o programa produz. Ele deve estar muito próximo de 100%. Na hora da escolha, não se iluda com percentuais de acerto de 95% ou 98,7%. Na prática, isso significa que de cada cem toques, erra-se de cinco a um carácter. Um datilógrafo com esse índice seria convidado a refazer o curso de datilografia. É possível aumentar esse índice, utilizando conjuntamente um editor de texto com dicionário em português.

Normalmente, um original costuma ter correções ou anotações feitas à mão. Nestes casos, contratar uma digitadora (humana) pode ser a escolha mais sensata. E, geralmente, quando o original está muito bom é porque ele foi digitado em um computador. Aí é só pedir o disquete com o texto original e passar para o seu Mac.

Tabelas, textos quebrados e fundos coloridos também são problemáticos. A MACMANIA, por exemplo, é um desafio

enorme para os programas OCR. Provavelmente, você levaria mais tempo corrigindo os erros do que redigindo toda a revista.

Com tantos problemas, qual seria então o usuário ideal? Simples: se a tarefa básica for colocar todo trabalho que você fez antes da era da informática ou transcrever um livro ou notícia publicada em jornal ou revista (desde que com tipos de letras conservadores), o OCR é uma boa pedida.

O melhor momento para se comprar um OCR é no ato da compra de um scanner. A maioria dos fabricantes de scanners oferece *bundles* opcionais com programas de OCR, com generosos descontos no preço dos softwares, que não são nada baratos. Fuja dos scanners de mão. Apesar de serem uma alternativa barata, eles têm uma má qualidade de escaneamento e vêm em conjunto com softwares que deixam bastante a desejar.

Um software que se sobressai entre a maioria dos programas de OCR (tanto em qualidade quanto em preço) é o OmniPage Pro 2.1 (US\$ 995), da Caere. É seguramente um dos que apresenta melhor taxa de reconhecimento e facilidade de uso. Sua capacidade para “aprender” a reconhecer caracteres diferentes toma um pouco de tempo no início, mas depois aumenta consideravelmente sua eficiência. €

## DICAS PARA UM BOM OCR

**Cubra o scanner** - Qualquer fiapo de luz que entre durante o processo de escaneamento pode comprometer seu OCR. Isso pode ocorrer se você está escaneando livros grossos. Cubra o scanner com um pano grosso ou xeroque as páginas do livro e escaneie a cópia.

**Ligue os pontos** - Fotocopiar páginas impressas em matriciais pode melhorar sua compreensão pelo OCR. Os pontos que formam as letras ficam maiores e mais próximos.

**Formato ideal** - Texto em corpo 10, fontes sem serifa (helvética, de preferência), poucas colunas e nada de rabiscos ou desenhos que atrapalhem o OCR. Escaneie apenas as partes a serem interpretadas pelo OCR. Deixe de fora logos, assinaturas, fotos e ilustrações que só causarão “ruídos” em seu OCR.

**Resolução** - A resolução ideal para um bom OCR é 200 dpi. Aumentar a resolução só traz resultados se o texto estiver em corpo muito pequeno. Alguns softwares não aceitam texto escaneado em resolução maior que 300 dpi.

8-) eu uso óculos :-( eu tenho bigode 8-) eu sou uma garotinha :-)-8 eu sou uma garotona :-Q eu fumo :-? eu fumo cachimbo