

Cartilha de Digitalização

Resumo

Configuração	Valor
Resolução	150 dpi
Profundidade de cor	Preto e branco
Formato do Arquivo	PDF/A
OCR	Durante ou após a digitalização

Configuração do Scanner

Quando você vai usar um scanner para digitalizar alguma foto e/ou documento, é preciso fazer alguns ajustes iniciais no software que acompanha o equipamento. Este software é o programa que gerencia as funções do scanner e pode ser acionado através de um editor de imagens, ou diretamente do atalho que o executa. Ele vem com o equipamento que você comprou e deve ser instalado em seu computador segundo as indicações de cada fabricante. Cabe ressaltar que a alteração de qualquer uma das definições a seguir afeta o tamanho do arquivo gerado após a digitalização. Portanto, para conseguir um documento apropriado para uso no Processo Eletrônico, deve-se proceder os ajustes de forma adequada. Normalmente, são necessários apenas três ajustes principais:

Resolução

O scanner "quebra" a imagem e armazena suas formas e cores em pequenos pontos chamados pixels. O valor da resolução define a quantidade de pontos por polegada (em inglês, dots per inch, ou dpi) a serem capturados. Assim, ao dizer que uma imagem tem 300 dpi, queremos dizer que em cada polegada da imagem há 300 pontos. Quanto maior este número, mais definida (e maior) será a imagem resultante no computador. Recomendamos o uso da resolução de **150 DPI**.

Profundidade de cores ou qualidade da cor

Basicamente, existem três opções: colorido, escala de cinza e preto-e-branco, sendo que uma imagem:

- Preto-e-branco ocupa 1 bit para cada pixel;
- 256 cores ou escala de cinzas ocupa 8 bits para cada pixel;
- 65.536 cores ou colorido (RGB) ocupa 16 bits para cada pixel;
- 16,8 milhões de cores ou True Color ocupa 24 bits para cada pixel.

Quanto maior a profundidade de cores (bits) maior a quantidade de informações capturada pelo scanner, e portanto, maior a similaridade entre a cópia e o original, no entanto, o arquivo gerado também será maior. Recomendamos o uso da profundidade de cor **Preto-e-branco**. Em caso excepcional, pode-se utilizar outra configuração de profundidade de cores quando houver a necessidade de informação em colorido.

Formato do Arquivo

O formato em que você salva seus documentos digitalizados também é uma decisão importante a ser tomada. Entre os mais comuns estão BMP, JPG, TIF, DOC, XLS e PDF. Cada um desses formatos possui suas vantagens e desvantagens, sendo que a escolha de um ou outro está diretamente relacionada com a utilização posterior do arquivo final, pois um mesmo documento ficará com tamanhos diferentes se gravado em formatos diversos. Para uso no processo eletrônico do TCE-PB, essa escolha deve ser pelo formato **PDF**.

Tornar o arquivo pesquisável (OCR)

A tecnologia **Optical Character Recognition (OCR)** reconhece caracteres a partir de um arquivo de imagem ou mapa de bits. Através do OCR é possível digitalizar uma folha de texto impresso e obter um arquivo de texto editável. É necessário que todos os arquivos PDF encaminhados ao TCE-PB possuam conteúdo pesquisável. É possível tornar um documento digital pesquisável em dois momentos:

- **No momento da digitalização:** nesse caso, deve-se marcar a opção de passar o OCR na hora da digitalização. Deve-se observar que nem todos os scanners possuem programas com esse recurso.
- **Após a digitalização:** com documento já digitalizado, pega-se o arquivo PDF e o processa em uma ferramenta de OCR específica para essa função. Assim, o arquivo PDF originalmente não pesquisável será transformado em um com conteúdo pesquisável. Sugestão de Ferramentas de OCR (OmniPage, TextBridge, Adobe Capture, Abbyy FineReader, FreeOCR, Online-OCR, WonderShare PDF Editor).

Conferência da Digitalização

Para assegurar que o documento foi corretamente digitalizado, é imprescindível que o digitalizador faça uma conferência dos seguintes pontos:

1. Conferência do número de páginas digitalizadas
 1. Se estiver sendo digitalizada apenas a frente das páginas, observar se o número da última página é o mesmo marcado acima da folha no aplicativo digitalizador. Caso haja alguma divergência, devem-se conferir as folhas. Exemplo: se foram digitalizadas 10 páginas, mas o número da última página é 11, provavelmente alguma página deixou de ser digitalizada.
 2. Se estiver sendo digitalizados frente e verso das páginas ou tons de cinza, a melhor opção é se passar as folhas verificando numeração.
 3. Eliminar as folhas em branco manualmente.
2. Rotacionar as folhas, caso necessário, de modo que o texto fique posicionado corretamente para a leitura.
3. Conferência de legibilidade: Na tela, ver a imagem em escala 1:1 - 100 %, verificar a legibilidade do documento;
4. Imprimir uma folha do documento digitalizado e fazer exame a "olho nu", da cópia impressa, para avaliar a qualidade da captura digital comparando com o documento original;
5. Testar o sistema de busca, procurando-se uma palavra qualquer do texto, a fim de determinar se o OCR foi devidamente aplicado;

Exemplos

Aplicar OCR em arquivo PDF - Adobe Acrobat Professional v. 9.0

Aplicar OCR em arquivo PDF - Abbyy Reader

Essa cartilha foi elaborada utilizando como referência o conteúdo dos documentos listados abaixo:

- Roteiro para Digitalização de Imagens
- Tutorial de digitalização dirigido ao uso do processo eletrônico
- Guia de Digitalização de Procedimentos submetidos à CGU/MG/AGU
- Manual de Boas Práticas (Grupo de Trabalho 6, Minerva)
- Guia para digitalização de documentos Versão 2.0 (Embrapa)
- Recomendações para digitalização de documentos arquivísticos permanentes (CONARQ)

