

MEMENTO PARA DIGITALIZAÇÃO DE DOCUMENTOS PDFs PESQUISÁVEIS E HACHURADOS

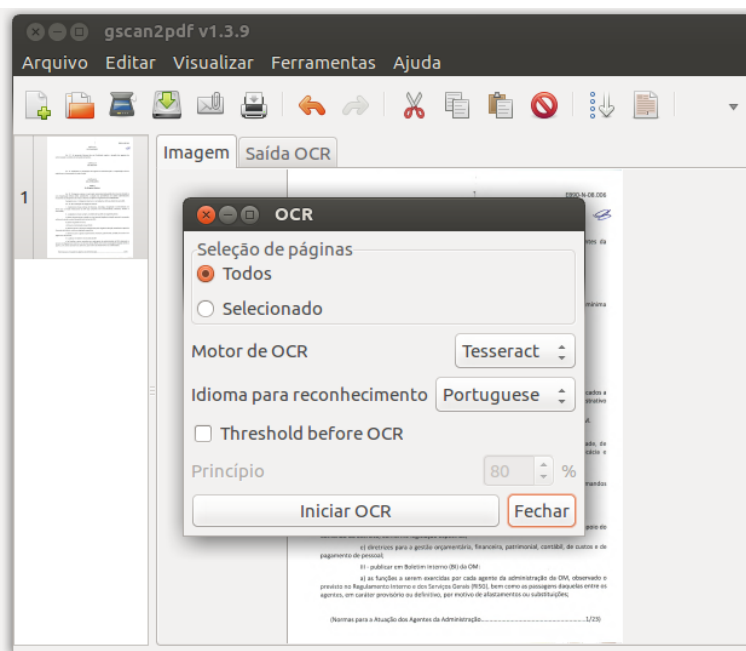
(Elaborado pelo 3º CGCFEx – versão 2.0)

1. Digitalizar (“scanear”) documentos pesquisáveis (OCR)

A maioria dos *scanners* oferece a opção de digitalizar documentos com o recurso **OCR – Optical Character Recognition**, recurso que possibilita a geração de documentos **PDF pesquisáveis**. Caso o equipamento disponível não possua esse recurso, basta realizar a digitalização para PDF sem OCR e utilizar um programa separado que realize o processamento de OCR no arquivo PDF. Sugere-se utilizar o programa de *Software Livre “gscan2pdf” do Linux (Ubuntu)* para realizar essa tarefa, conforme o passo a passo a seguir. Para instalar o programa, digite no terminal do *Linux* (caso seu usuário não possua permissão para instalar programas, solicite o apoio da Seção de Informática da sua OM):

```
# sudo apt install tesseract-ocr tesseract-ocr-por gscan2pdf
```

Concluída a instalação, execute o programa *gscan2pdf* através do menu principal do Sistema Operacional: *Aplicativos → Gráficos → gscan2pdf*. Abra o arquivo PDF desejado pelo programa (*Arquivo → Abrir*). A seguir, execute o reconhecimento de caracteres através do menu *Ferramentas → OCR*. Na janela do OCR, selecione “*Portuguese*” como o idioma para reconhecimento e clique no botão “*Iniciar OCR*”:

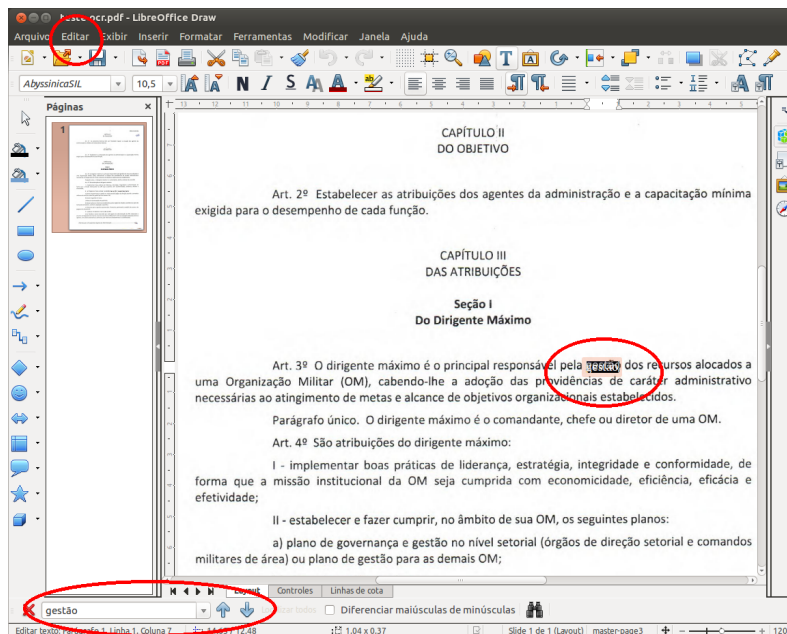


Após alguns segundos (dependendo da quantidade de páginas e do tamanho do arquivo sendo processado) o programa conclui a operação. Salve o arquivo PDF (agora “*pesquisável*”) através do menu *Arquivo → Salvar*.

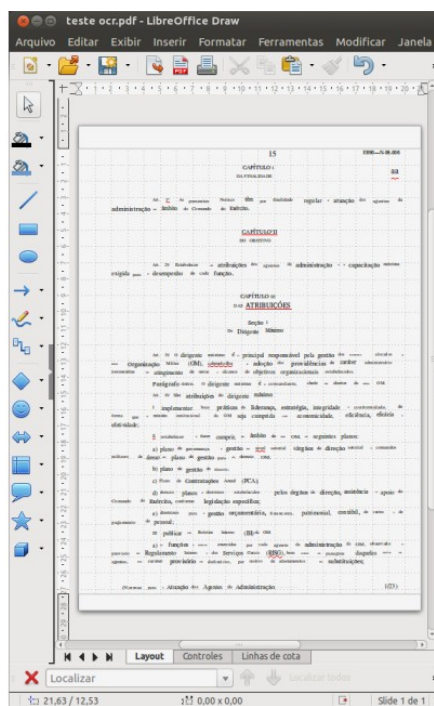
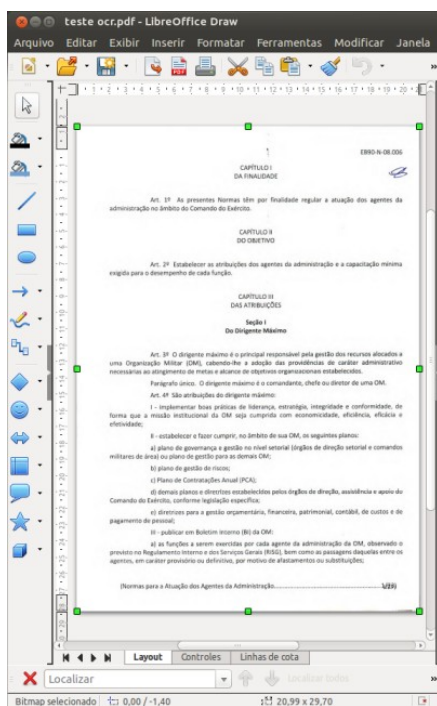
2. Hachurar Documentos PDF

Utilize o programa **LibreOffice Draw** para hachurar e salvar documentos PDF (caso não possua a suíte de aplicativos **LibreOffice**, solicite a instalação com a Seção de Informática de sua OM). Abra o arquivo PDF pelo LibreOffice através do menu **Arquivo** → **Abrir** ou pelo botão direito do navegador de arquivos do sistema operacional (Windows ou Linux) e selecione “**Abrir com** → **LibreOffice**”.

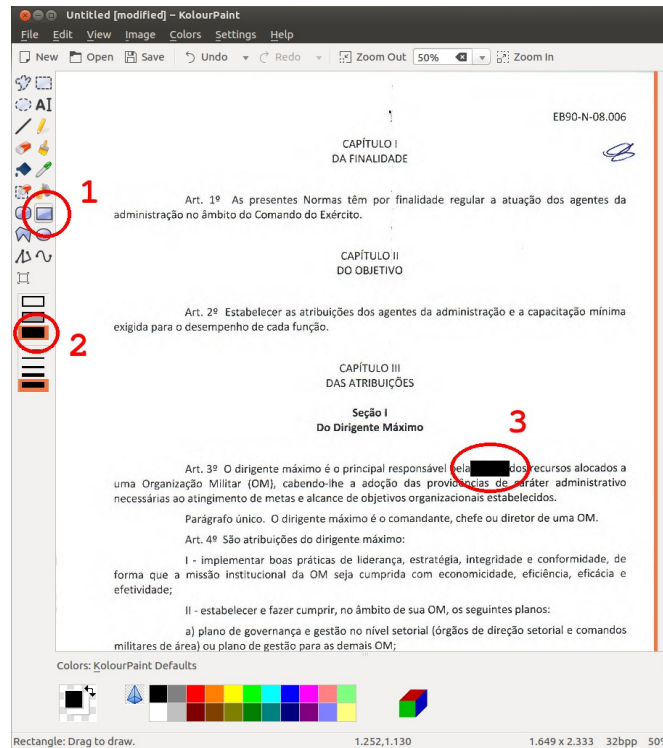
Encontre o texto a ser hachurado utilizando o menu “**Editar** → **Localizar...**” (ou **Ctrl+F**). O **LibreOffice** automaticamente seleciona o texto pesquisado; pressione a tecla “**Del**” ou “**Delete**” no teclado para apagar o texto selecionado. Este procedimento apagou o texto “*pesquisável*” a ser hachurado:



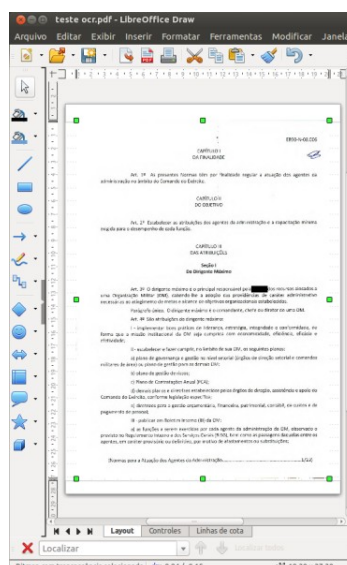
Agora é necessário, ainda, apagar a “imagem” desse texto no documento. Para tanto, clique com o botão direito do *mouse* sobre a página e selecione a opção **recortar**; a imagem da página digitalizada vai sumir, ficando visível apenas o texto OCR:



Para efetivamente remover a informação desejada da imagem da página, será necessário utilizar um outro programa editor de imagens. Sugerimos utilizar o *PaintBrush* no *Windows* ou *KolourPaint* no *Linux* (caso não esse programa no seu computador, solicite a instalação com a Seção de Informática de sua OM). No exemplo a seguir será utilizado o *KolourPaint*, mas o procedimento no *PaintBrush* é o mesmo. O primeiro passo é colar a imagem recortada do LibreOffice no KolourPaint, selecionando no menu “*Edit → Paste (Ctrl + V)*”. A seguir, selecione a ferramenta “*retângulo*” na barra de menu vertical esquerda (1), clique sobre o retângulo preenchido (2) e desenhe a forma geométrica que servirá de hachura para a “imagem do texto” desejado (3), conforme a figura abaixo:



Agora é necessário copiar e colar imagem da página de volta no *LibreOffice*: no menu do *KolourPaint*, selecione “*Edit → Select All (Ctrl + A)*” e, em seguida, “*Edit → Copy (Ctrl + C)*”; finalmente, no *LibreOffice*, selecione “*Editar → Colar (Ctrl + V)*”:



Repita o procedimento de localizar, apagar e hachurar para todas as ocorrências do texto desejado. Por fim, basta salvar esse arquivo editado, através do menu *Arquivo → Exportar como PDF* no *LibreOffice*.