

Glossário

- **Digitalização:** Processo de conversão de um documento para o formato digital, por meio de dispositivo apropriado, chamado digitalizador;
- **Digitalizador:** Periférico de entrada que digitaliza imagens, fotos e textos impressos e transfere para o computador, um processo inverso ao da impressora.
- **Documento digital:** Informação registrada, codificada em dígitos binários, acessível e interpretável por meio de sistema computacional.
- **Documento eletrônico:** Informação registrada, codificada em forma analógica ou em dígitos binários, acessível e interpretável por meio de equipamento eletrônico.
- **Formato de arquivo:** Especificação de regras e padrões descritos, formalmente, para interpretação dos bits constituintes de um arquivo digital. Os formatos de arquivo podem ser:
 - a. **aberto**, quando as especificações são públicas (p. ex.: .xml, .html, .odf, .rtf, .txt e .png);
 - b. **fechado**, quando as especificações não são divulgadas pelo proprietário (p. ex.: .doc);
 - c. **proprietário**, quando as especificações são definidas por uma organização que mantém seus direitos, sendo seu uso gratuito ou não (p. ex.: .pdf, .jpeg, .doc e .gif);
 - d. **padronizado**, quando as especificações são produzidas por um organismo de normalização, sendo os formatos abertos e não proprietários (p. ex.: .xml, .pdf/A).
- **Optical Character Recognition (OCR):** Tecnologia para reconhecer 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. Sigla para Reconhecimento Ótico de Caracteres, tecnologia usada em documentos digitalizados para identificar texto e possibilitar a pesquisa no documento. Dizemos que um arquivo possui OCR quando o texto dele pode ser pesquisado. Se o documento foi digitalizado como imagem é necessário utilizar uma ferramenta com recursos de OCR (como o Adobe Acrobat Professional) para reconhecer o texto e tornar o arquivo apto para a pesquisa textual.
- **Portable Document Format (PDF):** Formato de arquivo, desenvolvido pela Adobe Systems, para representar documentos de maneira independente do aplicativo, do hardware e do sistema operacional usados para criá-los. Os padrões PDF/X, PDF/E e PDF/A são definidos pelo ISSO (International Organization for Standardization – Organização Internacional para Padronização) Padrões PDF/X se aplicam à troca de conteúdo gráfico; padrões PDF/E se aplicam à troca interativa de documentos de engenharia; padrões PDF/A2 se aplicam ao arquivamento de longo prazo de documentos eletrônicos.