

SEÇÃO 3

AS MÚLTIPLAS FACES DO DOCUMENTO DIGITAL

MÓDULO 7

Gestão e preservação de documentos digitais

SEÇÃO 3

As múltiplas faces do documento digital

Adaptação do Arquivo Nacional da Costa Rica

Versão 1, 2024

Este curso foi traduzido e adaptado pela Direção Geral do Arquivo Nacional da Costa Rica, em colaboração com a Seção de Arquivologia da Universidade da Costa Rica, a partir do material original de 2011 da Associação Internacional de Arquivos Francófonos, disponível online no Portal Internacional Arquivístico Francófono. Esclarece-se que podem existir variações em relação ao conteúdo original. Para acessar o material em francês, visite <https://www.piaf-archives.org/se-former/module-7-gestion-et-archivage-des-documents-numeriques>.



ARCHIVO NACIONAL
COSTA RICA



UNIVERSIDAD DE
COSTA RICA

Conteúdo

Capítulo 1. Objetivo da seção	5
Capítulo 2. Conceitos básicos sobre o documento digital	6
2.1. O sistema binário	6
2.2. Bits e bytes	7
2.3. Unidades e ordens de grandeza	7
2.4. A representação hexadecimal	8
2.5. Primeiras noções de codificação	10
2.6. Dados, informação e documentos	11
Capítulo 3. O computador e seus componentes	14
3.1. Arquitetura básica	14
3.2. Os componentes do núcleo do computador	15
3.3. Os periféricos	16
3.4. Meios de armazenamento	18
Capítulo 4. Analógico e digital	19
4.1. Duas formas de memória da informação	19
4.2. Vantagens e limitações da representação digital	20
Capítulo 5. Diferentes tipos de documentos digitais	21
5.1. O documento «nascido digital»	21
5.2. O documento digitalizado	21
5.3. O documento híbrido	21

Capítulo 6. O documento digital no computador	22
6.1. O arquivo.....	22
6.2. O diretório	23
6.3. Sistema operacional.....	24
Capítulo 7. O documento digital	25
7.1. As redes	25
7.2. Os protocolos.....	25
Capítulo 8. Compressão de dados	26
8.1. O procedimento de compressão.....	27
8.2. Compressão com perda ou sem perda.....	27
8.3. A descompressão	27
Capítulo 9. Principais problemas para arquivar documentos digitais	28
9.1. Armazenamento e conservação de bits.....	28
9.2. Extrair a informação de forma inteligível	29
9.3. Encontrar nosso documento em meio a milhões de outros.....	29
9.4. Dispor de elementos probatórios sobre a integridade e autenticidade do documento	30
9.5. Conclusões	30
Bibliografia.....	31

Capítulo 1. Objetivo da seção

Antes de chegar ao conteúdo desta seção com mais detalhes sobre as diferentes formas ou faces que o documento digital pode assumir, é importante:

- Escrever a natureza particular do documento digital,
- Dar alguns elementos terminológicos essenciais,
- Começar a desmistificar sua estrutura interna,
- Mostrar como nasce e circula dentro do computador,
- Entender como este documento se relaciona com os arquivos e diretórios que lidamos

diariamente. Também é mostrado como o documento circula pelas redes e em que condições.

Será falado sobre a compressão de dados, uma função muito útil para arquivos de grande tamanho que não podem circular facilmente devido às limitações técnicas das redes atuais.

Finalmente, será apresentada uma primeira síntese dos diferentes problemas que devem ser resolvidos para garantir a conservação a longo prazo de um documento digital.

O objetivo específico é:

- Entender como organiza-se a informação em um espaço digital.
- Compreender a realidade das redes de comunicação e as tecnologias de transferência da informação



GLOSSÁRIO

Analogia: A memorização pode ser feita por analogia com o fenômeno. Fala-se de dados analógicos:

- Um termômetro registrador traçará as mudanças de temperatura ao longo do tempo em um rolo de papel. Os dados analógicos terão um caráter contínuo.

Autenticidade: Um documento autêntico é um documento do qual pode-se

provar a) que é o que afirma ser,

b) foi efetivamente produzido ou recebido pela pessoa que afirma tê-lo produzido ou recebido, e c) foi produzido ou recebido no momento em que se afirma que foi produzido ou recebido.

Bit: O bit (para “bynari digit”) é a menor unidade de informação manipulável por um computador. Um bit serve para especificar se a corrente passa ou não. Assim, só pode ter o valor 0 ou 1.

Compressão: recodificação de dados digitais para reduzir suas dimensões e poder assim economizar espaço de armazenamento ou tempo de transmissão. (Barnard, A e Voutssas, J, 2014, p. 47).

Compressão com perda: método de compressão de arquivos de dados, imagens, som, etc., usado para reduzir seu espaço; com esta técnica Alcança-se um fator de compressão de arquivos muito alto, mas o documento descomprimido perde um pouco de qualidade em relação ao original, pois reduz a quantidade de informação contida nos dados. (Barnard, A e Voutssas, J, 2014, p. 47).

Compressão sem perda: Método de compressão de arquivos de dados, imagens, som, etc., usado para reduzir seu espaço; com esta técnica o documento descomprimido é idêntico ao original, mas o fator de compressão não é muito alto. (Barnard, A e Voutssas, J, 2014, p. 47).

Integridade: Propriedade de um documento na qual seu conteúdo e características de identificação permaneceram inalteráveis desde o momento da sua emissão, ou então, que, tendo sido alterados posteriormente, foram com o consentimento de todas as pessoas legitimadas. (Arquivo Nacional da Costa Rica, 2020)

O documento não sofreu nenhuma modificação não registrada.

Rede: sistema de comunicação de dados formado por equipamentos e programas apropriados. As redes são frequentemente classificadas de acordo com sua extensão geográfica: rede de área local, rede de área metropolitana ou rede de área ampla. Também são classificadas pelo tipo de protocolo utilizado para se comunicar (*HTTP, token-ring*, etc). (Barnard, A e Voutssas, J, 2014, p. 184).

Capítulo 2. Conceitos básicos sobre o documento digital

Hoje em dia, um computador se tornou algo comum. Seja para o trabalho ou para uso mais lúdico, a maioria das pessoas utiliza essa ferramenta tecnológica. Muitos até a usam diariamente. Mas conhecemos realmente essa ferramenta da qual já não podemos prescindir? Como funciona?

O computador é o elemento vital da nossa sociedade da informação. Mas será que realmente conhecemos essa ferramenta tão versátil e preciosa da qual não poderíamos prescindir?"

2.1. O sistema binário

Entre os seres humanos, a informação pode ser transmitida de diferentes maneiras. Para isso utilizamos diferentes formas de “linguagens”, como a palavra, a escrita e também o gesto. No entanto, para trocar informações com um computador, essas formas de linguagem não são suficientes: a informação que desejamos transmitir deve ser codificada para que seja compreendida pelo computador, pois o computador fala sua própria linguagem, baseada no que chamamos de “sistema binário”.

No final dos anos 30, Claude Shannon demonstrou que uma máquina podia manipular informação. Como? realizando operações lógicas com ajuda de “contatores” fechados para “verdadeiro” e abertos para “falso”, associando ao “verdadeiro” o número “1” e ao “falso” o número “0”. Uma analogia com o anterior é pensar em um interruptor ligado ou desligado. De fato, o sinal básico de um computador se materializa pela ausência ou presença de uma corrente elétrica durante um período de tempo determinado:

- Se passar uma corrente elétrica pelo cabo do seu rádio durante um segundo, ele vai acender; essa ação pode ser representada pelo número 1.
- Se não passar corrente elétrica pelo cabo durante o mesmo período de tempo, o rádio permanece desligado e essa informação será representada pelo número 0.

Essa é a base da linguagem com o que todas as informações contidas em um computador são armazenadas e utilizadas.

2.2. Bits e bytes

O bit (para “bynari digit”) é a menor unidade de informação manipulável por um computador. Um bit especifica se a corrente passa ou não. Portanto, só pode ter o valor 0 ou 1. Um bit, por si só, não significa muito, mas agrupados de oito em oito, os bits formam bytes (em inglês, “bytes”), que podem representar diferentes informações, como as letras do alfabeto e os números

Concretamente:

Com um bit, podem ser representados dois estados diferentes: 0 ou 1,

com dois bits, podem ser representados quatro estados diferentes (2×2): 00 01 10 11,

e assim sucessivamente até um byte (um bloco de oito bits) que pode representar 256 estados diferentes: $(2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2) = 256$.

Os bits também serão utilizados para representar números inteiros em 2 ou 4 bytes, ou números decimais em 4, 8 e, às vezes, 16 bytes.

2.3. Unidades e ordens de grandeza

É importante e útil conhecer as principais unidades de medida de volume de dados digitais. Tudo isso é tão importante quanto ter referências concretas sobre o que cada uma dessas unidades de medida permite armazenar.

Unidade de medida	Equivalência	Exemplos
1 Kilobyte (Ko)	210 bytes	1024 bytes = fotografia de baixa resolução
1 Megabyte (MB)	220 bytes 1024 KB = 1 048 576 bytes	2 megabytes: uma fotografia de alta resolução 5 megabytes: as obras completas de Shakespeare 10 megabytes: um minuto de som de alta fidelidade 100 megabytes: 1 metro de livros em uma estante 500 megabytes: 1 CD-ROM
1 Gigabyte (Go)	230 bytes 1024 MB = 1 073 741 824 bytes.	1 gigabyte: uma van cheia de livros 20 gigabytes: uma coleção das obras de Beethoven

1 Terabyte (To)	240 bytes 1024 GB = 1 099 511 627 776 bytes	10 terabytes: o conjunto de documentos impressos da Biblioteca do Congresso dos Estados Unidos 400 terabytes: a base de dados do Centro de Dados Climáticos dos Estados Unidos
1 Petabyte (Po)	250 bytes 1024 TB = 1 125 899 906 842 624 bytes	1 petabyte: 3 anos de imagens de observação da Terra pelo EOS (Earth Observation Satellite), 200 petabytes: a totalidade do material impresso sobre o planeta 2 petabytes: o conjunto das bibliotecas universitárias americanas
1 Exabyte (Eo)	260 bytes 1024 Po = 1 152 921 504 606 846 976 bytes	5 exabytes: volume total da informação gerada sobre o planeta em 2002, 5 exabytes: todas as palavras pronunciadas pelos homens desde que existem.
1 zettabyte	1,024 exabytes	Armazenar todo o tráfego de Internet mundial durante um ano poderia requerer vários zettabytes de armazenamento. Isso incluiria todos os dados transferidos através de redes, como sites, e-mails, vídeos, redes sociais, transações online, etc.
1 Yottabytes (YB).	1,024 zettabytes	Em termos práticos, poderia imaginar uma capacidade de armazenamento tão grande que pudesse conter toda a informação gerada pela humanidade ao longo dos anos. Outro exemplo poderia ser a capacidade de armazenar todos os dados gerados por dispositivos IoT (Internet das coisas) a nível mundial durante um período prolongado de tempo, incluindo sensores, dispositivos conectados e mais.

Unidades de medida dados digitais. Elaboração própria a partir do texto de PIAF.

Os elementos de comparação propostos no quadro (até exabyte) procedem de um relatório intitulado “How much information”, publicado online em 2003 pela Universidade de Berkeley.

2.4. A representação hexadecimal

O computador manipula bits, geralmente organizados em bytes.

Na continuação deste curso, será abordada a estrutura interna de um documento digital. Este documento consiste em uma ou mais séries de bits. Uma série de bits pode ter uma longitude e uma forma incompreensível.

**EXEMPLO**

```
011011110111001101101000011011110111000000100001000101011011000110001101101 110010101101
1100111010001110011001000000001100110010111000110001000000100000010010110000 011000110101
1011100111010001101111011101101000000000000001100100011000110000.....
```

Dado que essa representação binária é pouco prática para ser manipulada por seres humanos, utiliza-se a representação chamada hexadecimal, construída sobre um sistema de base 16, compatível com o sistema binário (base 2) dos computadores, que é muito mais conciso.

“A representação hexadecimal ou representação em base 16 é uma notação condensada dos números binários. Observando que $24 = 16$, é possível representar a metade de um byte binário utilizando um dos 16 símbolos do sistema hexadecimal. Neste sistema os dez primeiros símbolos são idênticos aos utilizados no sistema decimal: 0, 1, 2, 3, 4, 5, 6, 7, 8 e 9, e os seis últimos correspondem às primeiras letras do alfabeto latino: A, B, C, D, E y F, que valem respectivamente: 10, 11, 12, 13, 14 e 15 em base 10.”

Hexadecimal	Binário	Hexadecimal	Binário
0	0000	8	1001
1	0001	9	1000
2	0010	A	1010
3	0011	B	1011
4	0100	C	1100
5	0101	D	1101
6	0110	E	1110
7	0111	F	1111

A representação hexadecimal dos bits. Elaborada por PIAF

Assim, será mais fácil representar a letra Z com o código hexadecimal “5A” do que com o octeto “1011010”.

Voltaremos a utilizar a representação hexadecimal em várias partes deste módulo.

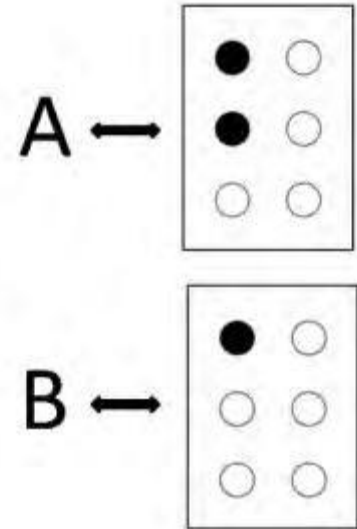
2.5. Primeiras noções de codificação

A codificação consiste em estabelecer uma correspondência entre um elemento com um primeiro nível de semântica e uma representação desse elemento em um sistema convencional.

Assim, a representação do fonema “A” do nosso alfabeto latino pelo símbolo gráfico “A” que utilizamos constitui uma codificação que é frequentemente percebido como implícito.

Exemplo. Sistema de escrita tátil Braille padrão.

Neste sistema, um caractere está representado pela combinação de 1 a 6 pontos em relevo, dispostos sobre uma matriz de 2 pontos de largura por 3 pontos de altura.



A representação das letras A e B no alfabeto Braille. Elaborada por PIAF

Exemplo. O alfabeto Morse é outro tipo de codificação:

O alfabeto Morse é um código que permite transmitir um texto por meio de uma série de impulsos curtos e longos. Essa codificação de caracteres atribui a cada letra, número e sinal de pontuação uma combinação única de sinais intermitentes. É um precursor das comunicações digitais que praticamente já não é utilizado. Aqui está a representação das letras A e B neste alfabeto.



A representação das letras A e B no código Morse. Elaborada por PIAF

É evidente que um arquivista que deve conservar documentos em código Braille ou código Morse também deve conservar a descrição do código em questão. Na representação binária, é o mesmo princípio.

Representação A e B

	Binaire	hexadécimal
A	1000001	41
B	1000010	42

Representação normalizada das letras A e B

sob a forma binária, assim como sua

representação hexadecimal. Elaborada por PIAF

2.6. Dados, informação e documentos

Para estabelecer uma terminologia comum a este módulo, propomos algumas definições essenciais. As definições dos termos “Informação” e “Dados” são as da norma ISO 14721: 2015 “Modelo de referência para um sistema aberto de arquivo”, norma essencial para o arquivo de documentos digitais, cujas características principais serão apresentadas na parte 5 deste módulo, dedicada em sua maior parte a esse modelo.

Um texto em uma folha de papel, um texto escrito em Braille, um arquivo proveniente de um programa informático de informática, uma imagem, uma tabela de números, um gráfico são contêineres de informação, ou seja, dados: quando transmitidos a um indivíduo, este poderá, sob certas condições, derivar uma informação que modifique um capital de conhecimentos pessoal ou coletivo.

A derivação da informação a partir de um dado pode requerer também a utilização de um aparelho.

O importante é que a informação seja independente da forma como é representada nos dados.

Também falaremos com frequência sobre o documento.



DEFINIÇÃO

Informação: A informação é definida como um conhecimento que pode ser trocado. Na prática, a informação é, portanto, um elemento de conhecimento que pode ser codificado para sua conservação, tratamento ou comunicação.

Dado: Uma representação de fatos, conceitos ou instruções de maneira formalizada, e adequada para sua comunicação, interpretação ou processamento por meios automáticos ou humanos (Departamento de Serviços Arquivísticos Externos, Arquivo Nacional da Costa Rica, 2020) Por exemplo: uma sequência de bits, uma tabela de números, os caracteres de uma página, uma gravação de áudio, etc. Um dado é um portador de uma informação ou de um fragmento de informação.

Documentos: informação criada, recebida e conservada como evidência e como ativo por uma organização ou indivíduo, no desenvolvimento das suas atividades ou em virtude das suas obrigações legais. (ISO, 2016)

O documento é, portanto, um “conjunto de informações registradas”, considerado como uma unidade que pode ser utilizável para efeitos de consulta ou como prova.

Documento digital: É um documento constituído de forma digital, seja de forma nativa/inicial, ou por digitalização de um documento analógico.

No campo digital, a associação entre o suporte e a informação que leva perde seu sentido. Ao longo do tempo, o documento poderá ser armazenado em suportes diferentes sem que isso altere sua natureza nem sua integridade.

Documento eletrônico: Qualquer manifestação com caráter representativo ou declarativo expressa ou transmitida por meio eletrônico ou informático será considerada juridicamente equivalente aos documentos que sejam concedidos, residam ou transmitam por meios físicos. O uso do suporte eletrônico para um documento determinado não dispensa, em nenhum caso, o cumprimento dos requisitos e formalidades que a lei exija para cada ato ou negócio jurídico em particular. (Departamento de Serviços Arquivísticos Externos, Arquivo Nacional da Costa Rica, 2020)

2.7. Um documento digital simples

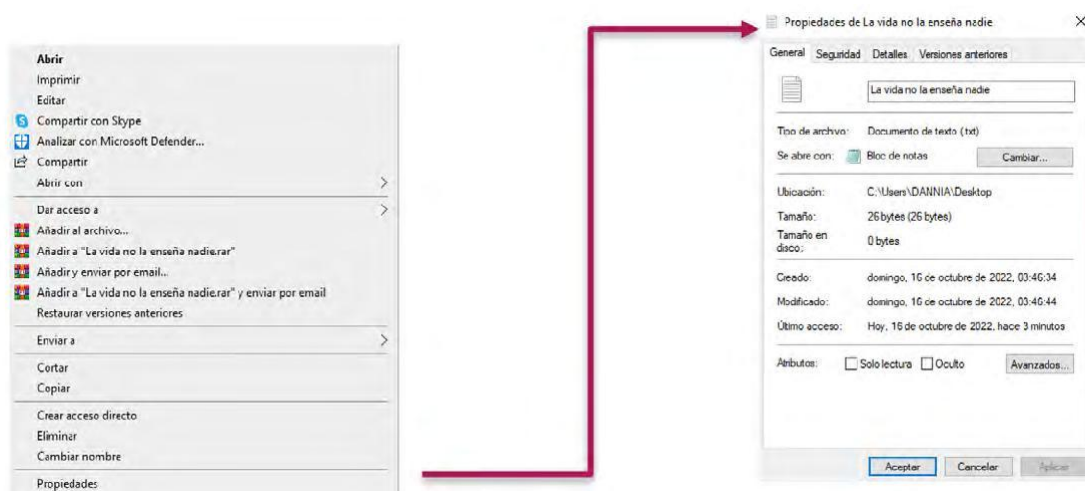
Criação de um **documento** pequeno: Vamos ver mais de perto o que pode conter um documento digital simples. Primeiramente, criamos para isso, com a ajuda de um editor de texto disponível em todos os computadores (Bloco de Notas, notepad...), um arquivo que contenha simplesmente a frase “A vida não ensina ninguém” (1), frase do conhecido escritor Gabriel García Márquez.

Em seguida, salvamos este arquivo em um diretório do nosso computador. O nome do arquivo será seguido de um ponto e da extensão “txt”, o que significa que se trata de um arquivo de texto simples, exemplo: A vida não ensina ninguém. txt (2)

Do ponto de vista da informação que contém, este arquivo constitui um documento muito elemental.

Além disso, em termos de codificação, **considera-se que o espaço entre duas palavras é um caractere autônomo**: o caractere “espaço” é portador de uma informação particular, a saber, que é necessário inserir um espaço para separar duas palavras. Levando em conta essa consideração, podemos contar que nosso arquivo contém 33 caracteres: 26 caracteres alfabéticos e sete caracteres “espaço”.

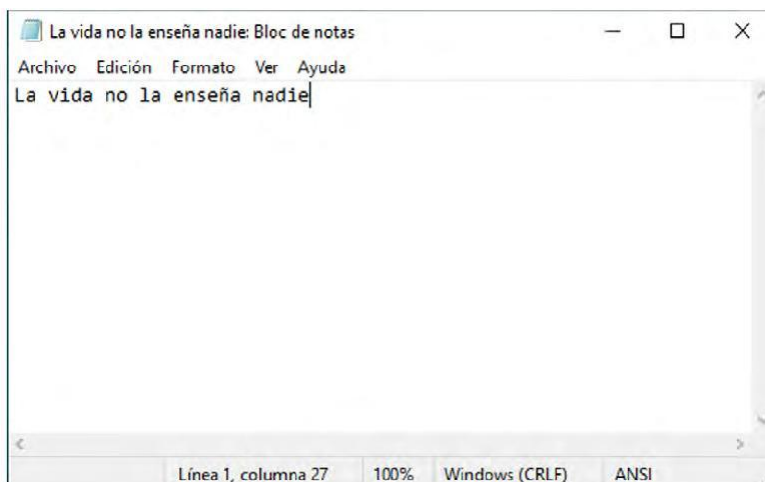
Ao apontar com o mouse para o nome do arquivo e clicar com o botão direito, teremos acesso a um menu contextual que nos permite acessar as propriedades do arquivo. Essas propriedades são resumidas em uma janela que se abre quando se aponta para 'propriedades' e se clica com o botão esquerdo do mouse, como mostrado nas figuras a seguir. 3.



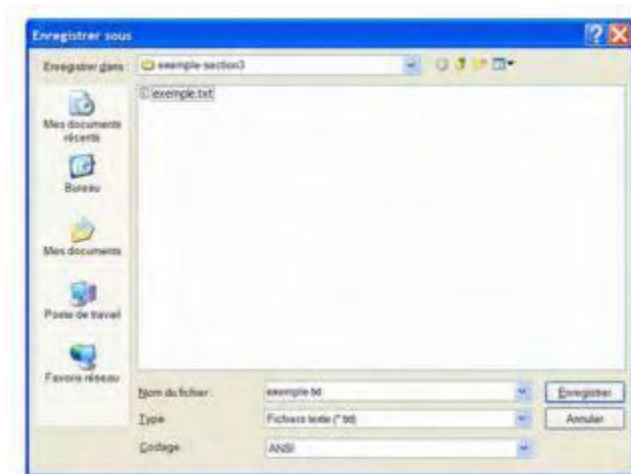
Menu textual e propriedades do processador de texto. Elaboração própria a partir de PIAF.

Nas propriedades, podemos observar que o tamanho do arquivo é de 26 bytes, o que é muito interessante, pois nos mostra que este arquivo é apenas um conjunto de bytes, e cada byte corresponde a um caractere.

É difícil imaginar algo mais simples.



A vida não ensina ninguém no Bloco de notas. Elaboração própria a partir de PIAF.



.txt. Elaboração própria a partir de PIAF.



COMPLEMENTO. Examinemos nosso pequeno documento.

O uso de um editor hexadecimal nos permite ver o que o arquivo realmente contém (e não apenas o que o editor de texto nos havia mostrado). Esse editor hexadecimal nos permite ver em que consiste realmente a continuação dos bytes contidos no arquivo:

4c 65 73 20 66 65 75 69 6c 6c 65 73 20 6d 6f 72 74 65 73 (em representação hexadecimal)

Ou talvez seja:

01001100 01100101 01110011 00100000 01100110 01100101 01110101 01101001 01100 01101100 01100
01100101 01110011 00100000 01101101 01101111 01110010 01110100 01100101 01110011

(em representação binária)

Dado que, na codificação normalizada utilizada, 4c corresponde à letra “L” (maiúscula), 65 corresponde à letra “e” minúscula, etc., podemos assim desvendar o conteúdo de nosso pequeno documento e constatar que não há nada de obscuro e mágico em tudo isso.

Capítulo 3. O computador e seus componentes

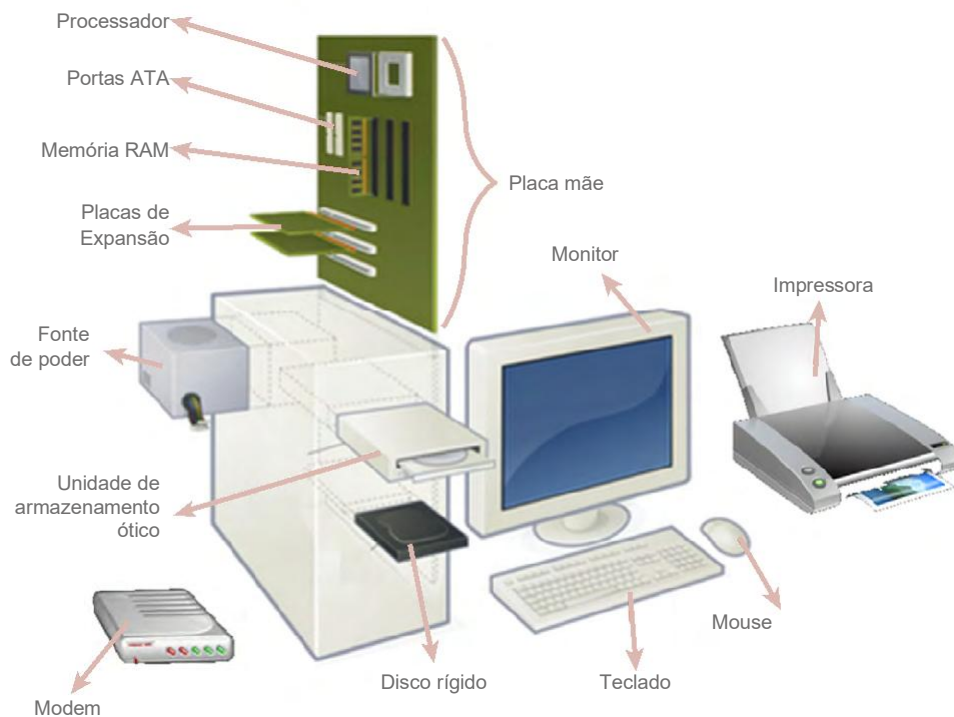
Agora que se entende um pouco melhor como a informação está codificada pelo computador, falta ver como ela está constituída dentro deste.

3.1. Arquitetura básica

Como vimos anteriormente, o computador faz apenas uma coisa: manipular números binários. A parte do computador que realiza todos esses cálculos é chamada de **processador ou microprocessador**. Mais adiante, veremos os componentes internos do computador com mais detalhes, mas principalmente, lembremos que o processador lê instruções na **memória** do computador (chamada memória RAM) e as executa.

Um programa é uma sequência de instruções. Tenha em mente que a memória do computador também contém dados sobre os quais o computador deve trabalhar.

O processador está conectado à memória por meio de dispositivos de comunicação chamados barramento de dados, que transportam, como o nome indica, todos os dados de um para o outro. Os “barramentos” são gerenciados por vários componentes, geralmente agrupados em um único chip: o chipset ou conjunto de chips. Todos esses componentes são projetados em torno de uma placa-mãe (um circuito impresso) que constitui o coração do computador.



O computador e suas partes

3.2. Os componentes do núcleo do computador

A seguir, são detalhados os componentes internos do computador; o importante não é lembrar de todos os elementos, mas compreender melhor seu funcionamento.

- **Processador ou microprocessador:** consiste em circuitos eletrônicos em miniatura agrupados em um só chip. O processador interpreta e executa as instruções. Atualmente, os processadores costumam ter múltiplos núcleos (dual-core, quad-core, etc.) para gerenciar tarefas de forma mais eficiente.
- **Barramento de dados:** transporta os dados entre os diferentes blocos funcionais e componentes de um computador. Um barramento que conecta apenas dois dispositivos é chamado de porta, os barramentos modernos são muito mais rápidos e eficientes, como o PCI Express, utilizado para a conexão de placas gráficas e outros dispositivos de expansão.
- **Chipset (conjunto de chips):** é o coração da placa-mãe e gerencia a troca de dados nos diferentes barramentos. Trata-se de um conjunto de circuitos integrados que geralmente são denominados «chips».
- **Placa-mãe:** É a placa principal de um computador que contém todos os componentes necessários para seu funcionamento. Entre eles estão a memória central, os barramentos e os conectores de expansão, as placas-mãe atuais frequentemente incluem recursos avançados, como conectividade Wi-Fi integrada, múltiplas portas USB, slots M.2 para unidades de armazenamento SSD ultrarrápidas, etc.
- **Relógio do sistema:** controla e sincroniza o microprocessador e os componentes associados. Sua frequência é expressa em gigahertz (GHz), pois os processadores modernos operam a velocidades muito mais altas.
- **A memória “cache”:** área de memória ultrarrápida integrada nos processadores, onde são armazenados as

instruções e dados que se repetem com mais frequência.

- **Conectores de expansão:** na placa-mãe, são espaços disponíveis para receber placas de usos diversos.
- **Entradas/saídas:** os computadores incluem portas seriais e paralelas para a conexão de dispositivos como impressoras, modems, entre outros. Agora, utilizam-se portas USB e Thunderbolt para a conexão de dispositivos. Também se tornou comum o uso de Bluetooth e Wi-Fi para a conectividade sem fio.
- **Memória de vídeo:** a memória, chamada VRAM, onde são armazenadas as imagens que são exibidas na tela.
- **Placas de vídeo e placas de som:** conectadas a um dispositivo de expansão, as placas de vídeo permitem descarregar o microprocessador dos cálculos da tela (ou seja, são responsáveis por tudo relacionado à exibição; é um circuito eletrônico independente que não está integrado à placa-mãe), enquanto as placas de som utilizam um driver para regular e controlar a entrada e/ou saída de áudio.

3.3. Os periféricos

Os dispositivos periféricos são dispositivos externos ao computador. Desempenham um papel fundamental na transferência de informação ao computador ou a partir dele.

Os dispositivos de entrada



Dispositivos de entrada

- **O teclado:** permite comunicar caracteres ou funções ao programa ou ao microprocessador, teclado e mouse têm aproximadamente as mesmas qualidades, digamos que são complementares.
- **O mouse:** permite selecionar e ativar o que é exibido na tela.
- **O scanner, a câmera de vídeo, a câmera digital ou a webcam, leitor de código de barras, caneta ótica, microfone** são também dispositivos de entrada, pois permitem a introdução de dados no computador.
- **Placa de assinatura digital:** permite a entrada de informações específicas por meio da assinatura digital, que é uma forma de autenticação e verificação de identidade em ambientes eletrônicos.
- **Leitor de cartões:** lê informações de uma fonte externa e as fornece ao computador.

Os dispositivos de saída:



Monitor



Impressora



Plotter



Caixas



de Som

Dispositivos de saída:

- **A tela/monitor:** permite a transmissão de informações visíveis ao usuário (texto, imagens, vídeo...).
- **Impressora e plotter:** permite a transferência de documentos do computador para um suporte de papel.
- **Caixas de som/alto-falantes e fones de ouvido:** permitem a saída de áudio do equipamento.

Os dispositivos de entrada e saída (Mistos):



Memória USB



Modem



Tela sensível ao toque



Disco rígido externo



DVD



CD

Dispositivos mistos

- **O disco rígido:** lugar de armazenamento principal dos dados. O tamanho de um disco rígido se conta em gigabytes (GB)
- **Reprodutor de CD-ROM e DVD:** dispositivo de reprodução que permite ler a informação contida em um CD-ROM ou DVD. No entanto, os novos equipamentos não incluem esse dispositivo devido à evolução tecnológica.
- **O leitor de memórias USB, o modem** (para as conexões à rede telefônica), a placa de rede (para as conexões a uma rede digital) também são dispositivos de entrada e saída.
- **Telas sensíveis ao toque:** os novos equipamentos estão implementados com essas ferramentas mais interativas.

3.4. Meios de armazenamento

Os documentos que são criados podem ser salvos pelo computador para serem utilizados posteriormente. Para isso, dispõem-se de diferentes meios de armazenamento chamados unidades ou dispositivos.

Esclarece-se que este é um meio de armazenamento de informações que pode ser utilizado na estação de trabalho habitual com o computador. Mais adiante, neste módulo, analisaremos a questão dos meios de armazenamento sob a ótica da conservação da informação a longo prazo.

Os principais meios de armazenamento utilizados são:

- **Disco rígido:** atualmente podem atingir uma capacidade de armazenamento de até 20 TB, no entanto, essa capacidade dependerá das características do equipamento que for escolhido, além disso, essa capacidade aumenta continuamente com os avanços da tecnologia. O sistema operacional e o software que são utilizados são armazenados no disco rígido, e não é necessário que ele esteja cheio para que o sistema operacional funcione corretamente.
- **Disco rígido externo:** a capacidade desses é variada e pode começar a partir de 250 GB, entretanto, quanto maior a capacidade de armazenamento, maior será o preço.
- **O Ultra HD Blu-Ray:** também chamado 4k Ultra HD é um disco ótico de alta resolução, que substituiu o Blu-Ray tradicional e possui uma capacidade de até 50 GB.
- (simples ou duplo camada). É utilizado principalmente para armazenar videogames em alta definição, mas na prática, permite armazenar qualquer tipo de documento digital.
- **Unidade USB:** sua capacidade varia hoje em dia somente algumas centenas de MB a 1 TB. É um meio de armazenamento de utilidade temporária para transferir dados de um computador a outro e tem a vantagem de ser fácil de transportar.
- **Nuvem:** sua capacidade depende do fornecedor do serviço e funciona como um meio de armazenamento remoto que permite acessar a informação de qualquer dispositivo conectado à Internet.

Neste ponto, recomenda-se realizar uma análise da informação que se pretende armazenar, pois dependendo dessa informação, será possível escolher o meio de armazenamento mais adequado de acordo com as necessidades. O anterior, devido ao fato de que nem todos os dispositivos possuem os mesmos níveis de segurança ao armazenar informações.

Capítulo 4. Analógico e digital

Dado que este módulo trata sobre documentos digitais, é importante explicar o que significa digital.

4.1. Duas formas de memória da informação

Chama-se sinal a representação física de uma informação. Na verdade, um sinal é a variação de uma magnitude (tensão, corrente elétrica, por exemplo) em função da duração. Os fenômenos que nos cercam, quando são quantificáveis, geralmente passam de um valor para outro sem interrupção: diz-se, então que esses fenômenos são contínuos. Para memorizar os valores desses fenômenos, é necessário registrá-los em um suporte físico que possa assumir valores contínuos.

O analógico e o digital são duas formas diferentes de memória da informação:

Analógico: cada vez mais, a memorização pode ser feita por analogia com o fenômeno. Fala-se de dados analógicos:

- A gravação de um disco de vinil reproduz a vibração do som no disco de plástico. O mesmo acontece com a gravação do som em um cilindro fonográfico, um termômetro registrador traçará as variações de temperatura ao longo do tempo em um rolo de papel.

Os dados analógicos terão um caráter contínuo.

Digital: a informação é transformada por um processo que a torna digital. Fala-se de dados digitais

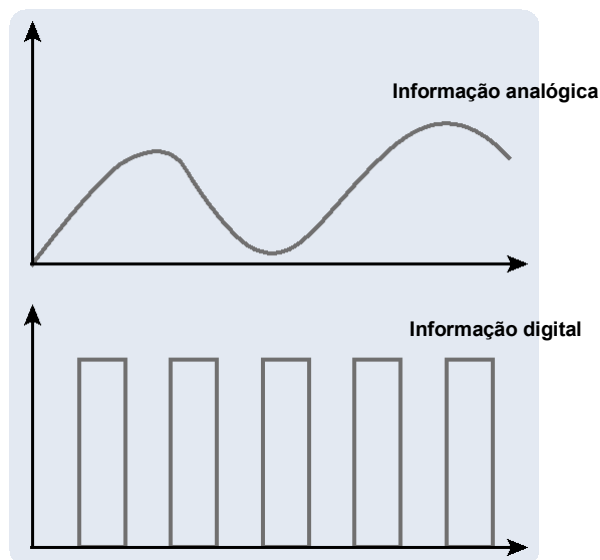
- a foto é escaneada (ou digitalizada pela câmera) para seu armazenamento,
- o som é mostrado e armazenado em um formato próprio da música (por exemplo, MP3),
- o registro das variações de temperatura será apresentado como uma série de pares de valores (temperatura 1, temperatura 1), (temperatura 2, temperatura 2), etc.

Neste caso, o sinal só pode assumir valores bem definidos, em números limitados, então fala-se de sinal digital. Um sinal digital se traduz assim por “0” e “1”; é, portanto, um número binário ou uma coleção de números binários.

Os dados digitais sempre se apresentam como uma sequência ou um conjunto de valores. Portanto, têm um caractere descontínuo, discreto.

O universo analógico existe há muitos séculos, se desenvolveu fortemente com o início da eletricidade, e seu correspondente digital surgiu mais recentemente com a era da informática.

Representação esquemática das duas formas de memória da informação.



Formas de memória. Elaboração própria a partir de PIAF.

4.2. Vantagens e limitações da representação digital

Uma das principais vantagens da tecnologia digital em relação à analógica é a possibilidade de armazenar, transportar e devolver os dados sem alterá-los. Também permite realizar facilmente transformações, tratamentos de todo tipo para criar nova informação.

De fato, o sinal analógico pode ser copiado (em uma fita magnética, em um disco de vinil, etc.), mas se degrada a cada etapa. Uma cópia de um sinal analógico é, portanto, de menor qualidade que a original, uma cópia de uma cópia é ainda pior e assim por diante, até que o sinal esteja completamente deteriorado e inutilizável.

Em comparação, a cópia digital é como um clone da original (chamado original múltiplo). O sinal, codificado de uma vez por todas, não se degrada e, em princípio, pode passar por vários suportes e canais diferentes sem perdas ou alterações de informação.



NOTA

A digitalização é o processo de transformação de um sinal analógico ou de objetos que estão no meio físico em um sinal digital.

No caso de um sinal de áudio digitalizado, este deverá ser convertido novamente em um sinal analógico para ser ouvido. No entanto, este é um processo do qual não nos damos conta, pois os dispositivos de música ou áudio são projetados para realizar essa conversão.

Capítulo 5. Diferentes tipos de documentos digitais

Para os efeitos deste capítulo, os documentos digitais poderão ser classificados em diferentes categorias de acordo com sua forma de criação.

5.1. O documento “nascido digital”

Trata-se de um documento criado diretamente no computador e que contém apenas elementos gerados pelo software que o criou, seja texto ou desenhos. Também falaremos de um documento digital nativo.

Neste caso fala-se de um documento nascido digitalmente.

As fotografias da câmera digital são da mesma natureza.

5.2. O documento digitalizado

A digitalização ou escaneamento pode ser realizada a partir de qualquer tipo de suporte, mas na maioria das vezes o documento “escaneado” ou “digitalizado” é originalmente um documento em papel (ou foto ou filme, etc.) “digitalizado” por meio de um digitalizador ou scanner. Este transforma a imagem de papel em uma grade muito fina, na qual cada elemento é chamado de pixel. A cada pixel corresponderá um valor na forma digital. Assim “digitalizado”, o documento em papel pode ser explorado por um software. Então será possível gerar cópias ou variantes sem que perca sua qualidade.

Da mesma forma, pode-se ‘digitalizar’ um documento sonoro ou audiovisual, com o objetivo de explorar mais amplamente seu conteúdo.

5.3. O documento híbrido

O documento híbrido é uma combinação dos dois primeiros.

Pode tratar-se, por exemplo, de um documento de texto criado no computador ao qual será associada uma imagem digitalizada. Também pode tratar-se de um documento nascido digital que foi digitalizado novamente a partir de uma versão impressa. Essa situação pode ocorrer quando o meio eletrônico é ilegível e não existem cópias, ou quando o formato do documento originalmente digital não permite sua conservação a longo prazo.



NOTA

Os documentos digitais podem classificar-se segundo critérios distintos de sua forma de criação. Seria possível distinguir, por exemplo, os documentos digitais “estáticos”, como uma imagem, e os documentos digitais “dinâmicos”, cuja visualização em uma tela utiliza vários programas ou fontes de informação. Esse é normalmente o caso de uma página web dinâmica, por exemplo, ou de um arquivo de planilha tipo “Excel” que contém dados de outros arquivos.

Dependendo da informação que se encontra nos documentos, é necessário avaliar a viabilidade de mudar o meio no qual se conserva, isso porque existem alguns elementos que permitem validar a informação em um meio, mas não funcionam para outro. Portanto, a fim de evitar que o documento perca validade, ele deve ser mantido no meio em que foi criado.

Este é o caso dos documentos digitais que possuem uma assinatura digital associada, pois a validade da assinatura não pode ser verificada se forem impressos, ou seja, se houver mudança de seu suporte original, já que não há como acessar as informações relacionadas ao certificado digital e outras propriedades significativas.

Capítulo 6. O documento digital no computador

O que acontece quando se cria um documento no computador, por exemplo, com um processador de texto? Onde é salvo e como ele se encontra?

6.1. O arquivo

No caso mais simples, o documento que você cria em seu computador com um software ou aplicativo é denominado arquivo. Pode ser um arquivo de texto ou outro tipo de arquivo (áudio, vídeo,).

As múltiplas faces do documento digital

Os arquivos podem ser armazenados em qualquer meio de armazenamento.

O tipo de arquivo ou ficheiro é reconhecido pela sua extensão. A extensão é o que vem após o ponto “.”.

Exemplos

- nome_do_arquivo.doc para um arquivo do tipo Microsoft Word (até Word 2003),
- nome_do_arquivo.txt para um arquivo criado com um editor de texto (Bloc note, textedit...)
- nome_do_arquivo.odt para um arquivo criado com OpenOffice Writer, o equivalente do Microsoft Word no mundo do software livre,
- nome_do_arquivo.docx para um arquivo criado com Microsoft Word 2007
- nome_do_arquivo.aiff para um arquivo de áudio criado em um computador Macintosh (Audio Interchange File Format)
- nome_do_arquivo.wav para um arquivo de áudio criado em um PC,
- Entre outros.

A extensão permite identificar o tipo de arquivo, que é criado por uma aplicação, ou seja, por um programa que cria e interpreta determinado tipo de arquivo.

Um arquivo “. doc” será criado com maior frequência pela aplicação Microsoft Word, mas outras aplicações como OpenOffice também são capazes de criar e ler arquivos com a extensão “.doc”.



COMPLEMENTO

Um arquivo pode ter uma estrutura interna muito simples ou complexa:

- Um arquivo do tipo “.txt” está constituído por uma série de caracteres como vimos na seção que aborda os conceitos básicos do documento digital,
- Os arquivos do tipo “.docx”, “.odt”, “.wav”, etc. têm estruturas internas mais complexas. Na realidade, são arquivos “contêineres”, ou seja, caixas nas quais podem ser encontrados vários ficheiros e diretórios.

6.2. O diretório

Um diretório ou “pasta” é o objeto informático que pode conter arquivos e outros diretórios, e que também pode ser armazenado em um meio (disco rígido, DVD, etc.).

Quando um diretório contém arquivos e outros diretórios, estes são chamados de subdiretórios; a pasta inicial que contém todos os diretórios é conhecida como “raiz” na maioria dos sistemas operacionais. Por último, chama-se caminho de acesso à localização de um arquivo ou diretório no computador.

Essa informação está localizada na barra de endereço do explorador do sistema operacional, que funciona da mesma forma que um navegador da web.

No Windows, acessa-se essa ferramenta pelo menu Iniciar, opção Programas, opção Acessórios. O Explorador de Windows é reconhecido pelo seu ícone, que é uma lupa dentro de uma pasta.

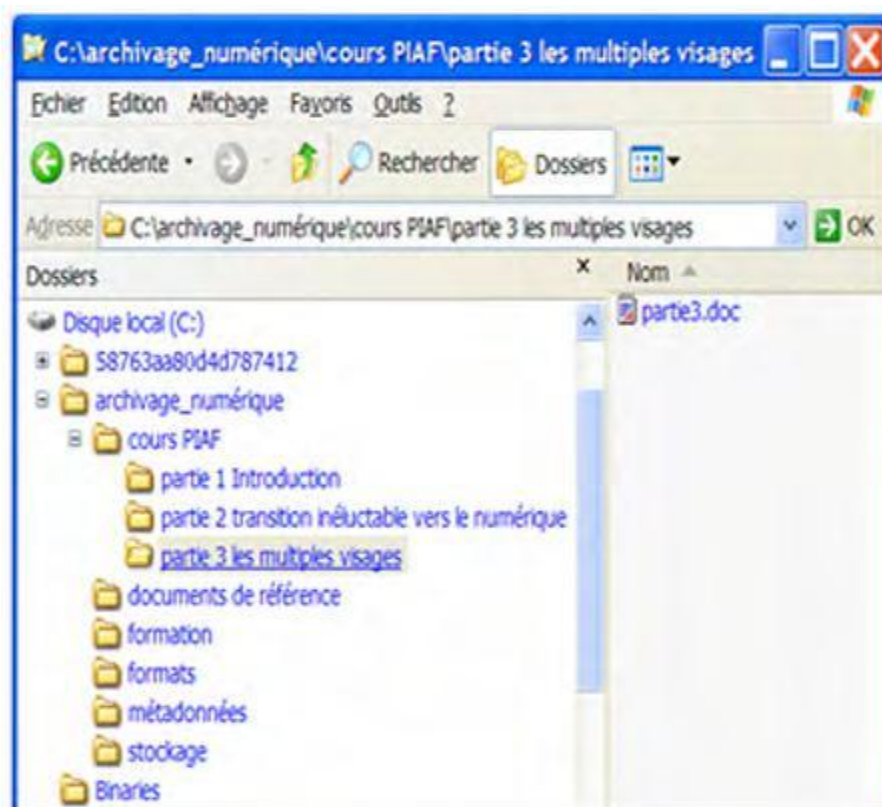
Por exemplo:

- No sistema operacional Windows, “C:” é a raiz da unidade C (o disco rígido) que contém diretórios, subdiretórios e arquivos. A seguinte estrutura mostra a posição de um documento no computador, ou seu caminho de acesso.
- Meio de armazenamento: disco rígido C:
- Diretório (ou pasta): “arquivo digital”
- Subdiretório (ou pasta): “curso PIAF”
- Subdiretório (ou pasta): “parte 3 as múltiplas faces”
- Arquivo: “parte3.doc”

O caminho de acesso ao arquivo é:

C: arquivo_digital.curso PIAF.parte 3 as múltiplas faces.parte3.doc

Representação, com internet Explorer, da árvore de diretórios descrita anteriormente.



Árvore de endereços. Elaborado por PIAF

6.3. Sistema operacional

O sistema operacional é o software principal do computador, que:"

- Fará os programas funcionarem,
- Permite gerenciar os arquivos e os diretórios estabelecendo vínculos entre o hardware, o usuário e as aplicações
- É utilizado para executar comandos básicos sobre as informações inseridas no computador, que são armazenadas lá ou que devem ser enviadas,
- Além disso, pode realizar outras tarefas, como a gestão de dispositivos (por exemplo, o controle de uma impressora) e a inicialização do sistema.

O mais conhecido dos sistemas operacionais é, sem dúvida, o Windows (com várias versões, sendo as últimas Windows 11, Windows Server 2019 e Windows 10 Mobile), mas há muitos outros, como o MacOS, que é exclusivo para os Macs, e o Unix em suas diferentes variantes (Linux, Solaris, HP-UX, entre outros).

Quando se fala de sistemas operacionais., fala-se frequentemente de sistemas multitarefa ou também de sistemas multiprocessadores.

Um sistema multitarefas é um sistema que permite compartilhar o tempo do processador para múltiplos programas.

Os programas parecem ser executados simultaneamente: portanto, é possível trabalhar com várias aplicações ao mesmo tempo.

(por exemplo, processamento de textos, planilhas e software de criação de páginas web).

Um sistema multiprocessador é um sistema multitarefa que permite organizar a execução simultânea de várias aplicações nos distintos processadores do sistema.

Esses podem ser estruturados em torno de um processador central ou ser independentes entre si e ter seu próprio sistema operacional. Eles se comunicarão entre si por meio de protocolos.



COMPLEMENTO

Vídeo: O que é um sistema operacional: sua história e evolução (MacOS, Windows, Android, iOS, Linux)

<https://www.youtube.com/watch?v=fPQCfy6FNE8>

Capítulo 7. O documento digital

O documento criado em um computador não precisa estar confinado! Na verdade, pode-se enviar a alguém ou levar consigo sem problemas.

As operações de transferência de documentos de um usuário para outro, de um computador para outro, são possíveis através de redes e protocolos de comunicação.

7.1. As redes

O documento digital pode viajar de um computador para outro através de meios de armazenamento (unidade USB, a nuvem), mas também através de redes.

As redes digitais permitem trocar informações de diferentes maneiras: por meio da transferência de arquivos, e-mail, troca de informações pessoais em tempo real (o “chat”), etc.

As redes geralmente se baseiam em infraestruturas de comunicação existentes, como a telefonia (com ou sem fio). As redes podem ser públicas: são criadas e administradas pelo Estado para suas próprias necessidades: administração, educação, investigação. Mas também podem ser privadas e próprias de uma empresa.

7.2. Os protocolos

Os computadores, assim como os humanos, elas devem falar o mesmo idioma para se entenderem.

Um protocolo é o conjunto de especificações que descrevem os convênios e regras que devem ser seguidas para permitir a comunicação entre duas máquinas.

Há vários protocolos, mas o mais conhecido é provavelmente o TCP/IP

- TCP (Transmission Control Protocol) para a “língua”,
- IP (Internet Protocol) para o método de encaminhamento de dados.
- TCP/IP é o protocolo utilizado por qualquer computador que utiliza Internet para se comunicar com outro.



COMPLEMENTO

De fato, há uma série de protocolos especializados que se ajustam a TCP/IP. Alguns dos mais importantes são os seguintes:

- FTP (File Transfer Protocol) para a transferência de arquivos
- HTTP (Protocolo de transferência de hipertexto) está relacionado com a navegação por Internet
- TELNET permite ao usuário acessar um computador host remoto e utilizar os recursos disponíveis.
- SMTP (Simple Mail Transfer Protocol) é o protocolo utilizado para o e-mail.
- SFTP (Secure File Transfer Protocol) y HTTPS (Hypertext Transfer Protocol Secure) são utilizados para transferências seguras: documentos confidenciais, transações bancárias.

Capítulo 8. Compressão de dados

É possível comprimir um arquivo para reduzir seu tamanho. Mas, por que reduzir o tamanho de um arquivo?

Existem duas razões principais:

- Facilitar a transferência para uma rede. As redes têm capacidades de transferência limitadas (essa capacidade será frequentemente expressa em Megabits por segundo. Quanto menor for o arquivo, mais curto será o tempo de transferência. Além disso, muitas organizações impõem uma limitação no tamanho dos documentos que podem ser anexados a um e-mail como arquivo. Também neste caso pode ser interessante comprimir o arquivo.
- Armazená-lo em um meio cuja capacidade de armazenamento é limitada. Assim, pode-se desejar, por exemplo, reduzir um arquivo cujo tamanho exceda a capacidade de uma unidade USB. É possível que deseje economizar espaço disponível no disco rígido. O sistema operacional Windows oferece, por exemplo, comprimir arquivos que não foram utilizados por muito tempo.



ATENÇÃO

A compressão com fins de conservação a longo prazo envolve vários riscos que convém analisar.

8.1. O procedimento de compressão

Existem muitas técnicas de compressão, mais ou menos eficientes e mais ou menos caras para colocar em funcionamento. O software de compressão reescreverá a informação de maneira mais concisa, por exemplo, poderá substituir os dados repetitivos do arquivo por um código.

Exemplo

Aqui está um caso muito simples:

Uma imagem com várias linhas de pixels da mesma cor vermelha (por exemplo, 250 pixels).

Em vez de guardar cada um desses pixels como um número para representar essa cor (o número 110, por exemplo), o software de compressão emite um código que significa: colocar aqui 250 vezes a cor nº 110.

Um código é muito mais compacto que 250 códigos.

Deste modo pode-se reduzir pela metade o tamanho dos arquivos, às vezes mais.

Geralmente, as técnicas de compressão utilizam métodos matemáticos que vão mais além deste curso.

8.2. Compressão com perda ou sem perda

A compressão sem perda ajuda a restaurar o arquivo original com uma exatidão absoluta. É isso que geralmente se utiliza para a maioria dos arquivos.

A compressão com perdas é geralmente utilizada para imagens, arquivos de som ou vídeos. É possível comprimir um documento de modo que não se percam alguns detalhes, frequentemente indetectáveis por um ser humano, e assim reduzir consideravelmente o tamanho do arquivo. Claro, a integridade do documento original já não é totalmente preservada.

Portanto, é possível arquivar dados respeitando totalmente a integridade de um documento e utilizar uma compressão com perdas para sua difusão, desde que o documento difundido atenda às necessidades dos usuários.

Exemplo

As imagens em formato PNG são comprimidas sem perda, enquanto as imagens no formato JPEG, frequentemente, são objeto de uma compressão com perdas

Existem muitos formatos de arquivo de compressão. Geralmente serão reconhecidos pelo sufixo do nome do arquivo:

7z, zip, ace, rar, etc.

8.3. A descompressão

A descompressão costuma ser muito fácil.

Exemplo

Com o software Win Zip, por exemplo, você só precisa abrir o arquivo clicando nele e depois extrair o arquivo, com a função “Extrair”. Então, é possível escolher a localização do arquivo a extrair, ou seja, onde se deseja colocá-lo.

Alguns programas, como WinRAR, permitem descomprimir arquivos em formato rar, mas também em zip, ace.



COMPLEMENTO

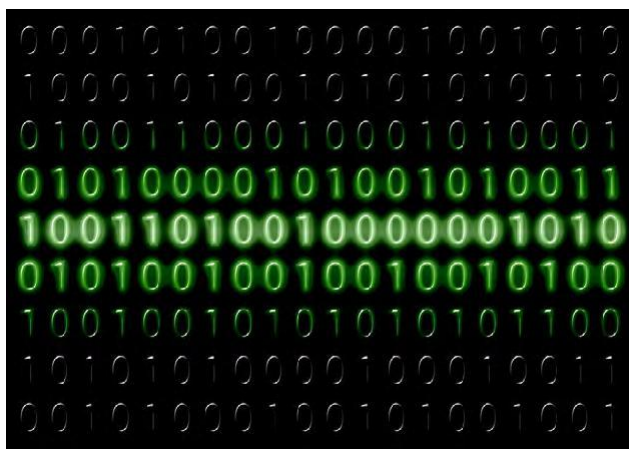
Vídeo:

<https://www.bilib.es/actualidad/blog/noticia/articulo/10-herramientas-digitales-claves-para-comprimir-y-descomprimir-archivos/>

Capítulo 9. Principais problemas para arquivar documentos digitais

No final desta seção sobre as múltiplas faces do documento digital, é necessário identificar os principais problemas que terão que ser resolvidos para a conservação a longo prazo dos documentos digitais. Esses problemas serão tratados detalhadamente mais adiante neste curso.

9.1. Armazenamento e conservação de bits



Documento digital Elaborado por PIAF

O documento digital é sempre apresentado na forma de uma ou mais sequências de bits.

O primeiro problema o que deverá ser resolvido é garantir a preservação e integridade dessas sequências de bits ao longo do tempo. Esse ponto será analisado na seção 6, dedicada ao armazenamento.

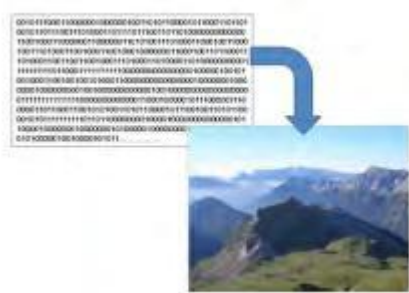
9.2. Extrair a informação de forma inteligível

Para entender esta seção, podem ser geradas as seguintes interrogações:

1. Como passar do documento em sua forma digital para uma informação inteligível que possa ser lida, entendida e interpretada?
2. Pode-se pensar que basta conhecer o formato do arquivo? A pergunta é muito mais complexa do que se poderia imaginar:
3. O formato é permanente ou desaparecerá dos ambientes técnicos a curto ou médio prazo? Há milhares de formatos de dados. Alguns desaparecem, outros são inventados. Os formatos proprietários não publicados estão vinculados ao software dos próprios proprietários,
4. Teremos os programas de computador que permitam ler os dados nesse formato dentro de 10 ou 20 anos?
5. A especificação do formato que se possui é completa, precisa, exata? Quem garante a conservação a longo prazo?
6. Os dados são realmente consistentes com o formato suposto? Foi verificado?

Além disso, a questão do formato resolverá apenas uma parte do problema. Para obter informações científicas, estatísticas econômicas e complexas, é necessária informação adicional para acessar realmente a semântica da informação.

Essas são as questões relacionadas com os formatos que são abordados na seção 7.



Uma pergunta central, transformar bits em um documento inteligível.

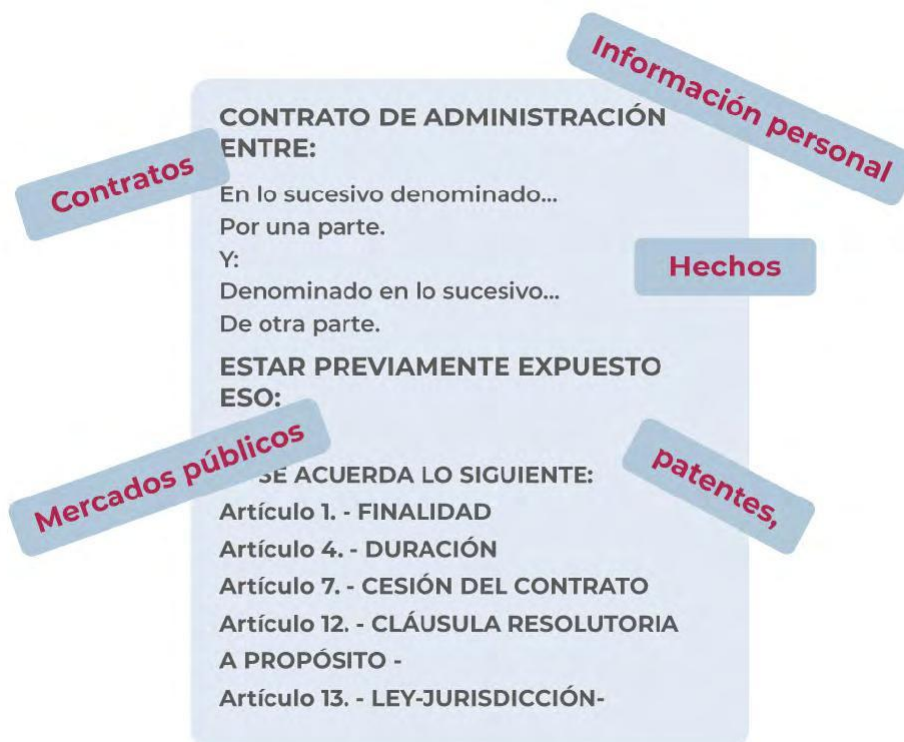
Dos bits ao conteúdo. Elaborado por PIAF

9.3. Encontrar nosso documento no meio de milhões de outros

Como avaliar a idoneidade de um documento para um determinado uso? Este será o objetivo dos metadados descritivos, que permitirão ao futuro usuário descobrir os documentos que correspondam às suas necessidades.

Os metadados também desempenharão um papel vital para ajudar o usuário a compreender e interpretar corretamente os documentos dos arquivos: é fundamental poder identificar os documentos de forma única e, sobretudo, duradoura, saber por que e em que condições esse documento foi elaborado, quais são as relações entre um documento específico e outros documentos arquivados, entre outros. Essas questões serão analisadas na seção 9, relativa aos metadados.

9.4. Dispor de elementos probatórios sobre a integridade e autenticidade do documento.



Que valor probatório tem o documento digital? Elaboração própria a partir de PIAF.

Pode-se confiar no documento digital?" Que elementos se tem para demonstrar sua integridade e autenticidade? Qual pode ser o valor deste documento diante de um tribunal, no contexto de um processo judicial?

Todos esses aspectos jurídicos serão tratados na seção 10.

9.5. Conclusões

A resolução desses quatro problemas essenciais: armazenamento, formatos, metadados e valor probatório não levanta apenas questões técnicas.

Ainda existem muitas necessidades em termos normativos e jurídicos, assim como os aspectos organizacionais; as estratégias também desempenharão um papel essencial. Por último, as questões de custo pesarão muito em todas as decisões.

Agora, você já está preparado para abordar todos esses temas com maior profundidade.

Bibliografia

- Archivo Nacional de Costa Rica. Departamento Servicios Archivísticos Externos (2020). Glosario único de términos, definiciones, conceptos y abreviaturas de las normas técnicas nacionales. https://www.archivonacional.go.cr/web/dsae/glosario_%20unico_terminos.pdf
- BANAT-BERGER F., HUC C., DUPLOUY L., *L'Archivage numérique à long terme, les débuts de la maturité?* (Primera obra de síntesis sobre el archivo digital en lengua francesa) Paris, La Documentation française, 2009
- BANAT-BERGER F., HUC C., Module 7. Gestion et archivage des documents numériques. Portail International Archivistique Francophone. 2011. <https://www.piaf-archives.org/se-former/module-7-gestion-et-archivage-des-documents-numeriques> (Se identifica en el texto como PIAF)
- Barnard, A y Voutssas, J (2014). Glosario de Preservación Archivística Digital Versión 4.0. Universidad Nacional Autónoma de México. https://iibi.unam.mx/archivistica/glosario_preservacion_archivistica_digital_v4.0.pdf
- Centro de Desarrollo de Competencias Digitales de Castilla La Mancha (2022). <https://www.bilib.es/actualidad/articulos-tecnologicos/post/noticia/10-herramientas-digitales-claves-para-comprimir-y-descomprimir-archivos/>
- ISO (2016) Información y documentación. Gestión de documentos. Parte 1 conceptos y principios. ISO 15489-1.

