

1  
Exercício : 1

\_\_\_\_\_ / \_\_\_\_\_ / \_\_\_\_\_

## Compressão de dados

A compressão de dados visa diminuir a redundância de dados. Logo apenas conjuntos de dados que tenham redundância podem ser comprimidos com vantagem. A compressão pode se dar em camadas: a. aplicativo: o aplicativo controla. Exemplo: uma tabela de cidades; b. do ambiente: o sistema operacional aplica. Para o aplicativo nenhuma mudança surge. Exemplo: PKZIP; c. Física: o hardware comprime. Exemplo: compressão feita pelo modem.

Os benefícios esperados são: a. diminuição de espaço e de tráfego. b. criptografia; c. melhoria de uso da CPU (já que os aplicativos em geral são IO bound); d. ultrapassar limites de mídias (exemplo um DVD). Certos arquivos são altamente propícios à compressão, por causa do fenômeno de "localidade de referência". Exemplos: música digital, vídeo...

Conceitos: Razão de compressão: tamanho descomprimido ÷ tamanho comprimido. Figura de mérito: o inverso da razão de compressão. Fração de redução: 1 - figura de mérito.

### Prática Localize um arquivo médio de texto:

Nome: \_\_\_\_\_. Tamanho: \_\_\_\_\_. Localize um compressor (ex: PKZIP) Faça a compressão rápida (pkzip: -es).

Tempo: \_\_\_\_\_, Tamanho: \_\_\_\_\_,

Razão: \_\_\_\_\_, Fração redução: \_\_\_\_\_.

Faça a compressão máxima (pkzip: -ex).

Tempo: \_\_\_\_\_, Tamanho: \_\_\_\_\_,

Razão: \_\_\_\_\_, Fração redução: \_\_\_\_\_.

Localize uma imagem BMP:

Nome: \_\_\_\_\_. Tamanho: \_\_\_\_\_.

Faça a compressão rápida (pkzip: -es).

Tempo: \_\_\_\_\_, Tamanho: \_\_\_\_\_,

Razão: \_\_\_\_\_, Fração redução: \_\_\_\_\_.

Faça a compressão máxima (pkzip: -ex).

Tempo: \_\_\_\_\_, Tamanho: \_\_\_\_\_,

Razão: \_\_\_\_\_, Fração redução: \_\_\_\_\_.

## Técnicas Usuais

**Código Morse** Só aparece aqui por razões históricas. O código morse é um sistema de representação de letras, números e sinais de pontuação através de um sinal codificado enviado intermitentemente. Foi desenvolvido por Samuel Morse e Alfred Vail em 1835, criadores do telégrafo elétrico (importante meio de comunicação à distância), dispositivo que utiliza correntes elétricas para emissão ou recepção de sinais. Veio na continuação do telégrafo visual em uso na França desde o século XVIII.

Uma mensagem codificada em Morse pode ser transmitida de várias maneiras em pulsos (ou tons) curtos e longos: \* pulsos elétricos transmitidos em um cabo; \* ondas mecânicas (pertubações sonoras); \* sinais visuais (luzes ascendendo e apagando); \* ondas eletromagnéticas (sinais de rádio);

Este sistema representa letras, números e sinais de pontuação apenas com uma sequência de pontos, traços, e espaços.

Com o desenvolvimento de tecnologias de comunicação mais avançadas, o uso do código morse é agora um pouco obsoleto, embora ainda seja empregado em algumas finalidades específicas, incluindo rádio faróis, e por CW (contínuas wave-ondas contínuas), operadores de radioamadorismo. Código morse é o único modo de modulação feito para ser facilmente compreendido por humanos sem ajuda de um computador, tornando-o apropriado para mandar dados digitais em canais de voz.

Os sinais são:

1. Sinal curto, ponto ou 'dit' (.)
2. Sinal longo, traço ou 'dah' (-)
3. Intervalo entre caracteres (entre pontos e traços)
4. Intervalo curto (entre letras)
5. Intervalo médio (entre palavras)
6. Intervalo longo (entre sentenças)

Como o comprimento variável de caracteres do código morse dificulta a adaptação à comunicação automatizada, o Morse foi amplamente substituído por formatos mais regulares, incluindo o Código Baudot e ASCII.

Para construir a tabela morse, usa-se a busca dicotômica na sequência, com ponto à esquerda e traço à direita.  
E,T,I,A,N,M,S,U,R,W,D,K,G,O,H,V,F,Ü,L,Ä,P,J,B,X,C,Y,Z,Q....

### Exemplo Traduza para o português:

...-1...-1.-12.1.-12. Resposta: \_\_\_\_\_.

**Comprimento de fileira** Escolhe-se um caractere pouco frequente (no nosso exemplo será Z). Sempre que um conjunto de mais de 3 caracteres se repetir, todos eles serão substituídos pelo conjunto Z+caráter a transmitir +quantidade de repetição.

No exemplo A.....DIVIDA.....E.DE.0000000033.REAIS, ficaria AZ.5DIVIDAZ.7E.DE.Z0833.REAIS

Observações: 1.naturalmente, se o caractere de contagem for um byte binário, o contador poderá chegar até 258, pois ele pode começar em 4, neste caso, reservando-se o '0' para a ocorrência simples do caractere.

Se por acaso, o caractere de controle existir no arquivo, ele pode aparecer na saída como duplicado, com tamanho 1.

Exemplo: A.....LUZ.E.BRANCA ficaria AZ.7LUZZ1.E.BRANCA

### Exemplo Comprima usando esta técnica:

AAABBBBBB000000574-REAIS. Resposta: \_\_\_\_\_.

### Exemplo Traduza para o português:

EU\~6QUE\R80\~0672. Resposta: \_\_\_\_\_.

**Substituição de padrões** A substituição de padrões é –em geral – um método de compressão de aplicativo, já que é dependente dos padrões da aplicação. É muito usada em compressão de códigos fonte (os padrões são as palavras reservadas) ou em grandes sistemas de dados comerciais (nomes de pessoas, de ruas, de cidades, de conjuntos de zeros, CEPs, etc).

Nestes casos há uma tabela com os padrões mais freqüentes (256, 65536,...) que são convertidos para 1 ou 2 bytes. Se o arquivo for binário, não há este limite, pode-se usar qualquer tamanho de tabela de padrões ( $2^9$ ,  $2^{10}$ ,  $2^{11}$ ...). Apenas há que se sinalizar a existência do padrão comprimido. Pode ser o sempre presente caractere de controle acima visto.

**Huffman** Este método já foi estudado lá em estruturas de dados. Continua o mesmo. A codificação de Huffman é um método de compressão que usa as probabilidades de ocorrência dos símbolos no conjunto de dados a ser comprimido para determinar códigos de tamanho variável para cada símbolo. Ele foi desenvolvido em 1952 por David A. Huffman que era, na época, estudante de doutorado no MIT, e foi publicado no artigo "A Method for the Construction of Minimum-Redundancy Codes".

Supondo o dicionário:

P0010000	U0100
V0010001	M0101
C001001	D0011
Z01100	E0000
N01101	R0001
G00101	0111
B1100	.100
I1101	A101
L0111	

**Exemplo** Descomprima:  
26 98 68 D8 B4 (38 b.s.) Resposta: \_\_\_\_\_.

 **Para você fazer** Traduza para português, as mensagens morse

- 1) .1...1-..1---1.-..1.-12
- 2) -.-.1....1.-1--.1.1..-12
- 3) -.-.1.-1-.1.1-1.-12

Comprima as frases a seguir, usando o método de comprimento de fileira. Use a \ como caractere de controle.

- 4) YYYYYYYYYXXXXXXXXXXXXXZZZZZZZDDDDDJ
- 5) 22----REAIS-----E-----00000012345

Descomprima as frases a seguir, usando o método do comprimento de fileira. O caractere de controle usado é a \.

- 6) CICERO-DIZIA\~7CAI\~5FORA\~5JA--
- 7) \A7-\B5-\C7-\6\G5H\~7

Descomprima:

- 8) 20 06 3E 67 90 38 bits significativos
- 9) CE F6 56 23 C0 36 bits significativos
- 10) AC 47 95 BA F0 38 bits significativos

RESPONDA NO VERSO