**Símbolos**

A - significado: letra (1a letra do alfabeto)

Associar símbolos a uma informação.

Quando passamos para o digital há uma convenção para que os dados e símbolos possam ser representados como "zeros" e "uns". Como, por exemplo, a tabela ASCII onde o A é representado por 01000001, o B 01000010, o C 01000011... posso também representar o A por 65, que é a conversão do binário para decimal, também podemos fazer o mesmo para hexadecimal. A questão é, um determinado dado pode ser representado de diversas maneiras.

Na compressão, como devemos representar um determinado dado de uma nova maneira, utilizando uma codificação diferente, onde a mesma informação possa ser representada por uma quantidade menor de bits que o requerido originalmente.

A compressão de dados serve para comprimir dados com aplicações para arquivamento: ocupar menos espaço. A transmissão também envolve compressão de dados, há uma compressão durante a transmissão, ele vai sendo transmitido e ao chegar no destino o mesmo é descomprimido e então gravado em disco. Há alguns protocolos que fazem esse tipo de conversão.

Compressão de dados se dá de duas maneiras distintas:

**Sem perdas** - significa que o arquivo onde foi realizado a compressão ao ser descomprimido retorna para seu tamanho/arquivo original. Há 100% de preservação da informação. Tentam aumentar a redução de bits, realizando uma análise estatística da ocorrência de coisas. Por exemplo, a letra A aparece muitas vezes durante um texto, portanto é melhor tentar representar a letra A com um número menor de bits, o que não ocorre com a letra W, por exemplo, que raramente aparece em um texto.

**Com perdas** - o processo de compressão é irreversível. Uma vez feita a compressão, não há como voltar para o arquivo original. São algoritmos baseados na percepção de dados: *jpeg*, *mpeg*.

Conceitos:

comprimento médio da palavra de código (fórmula do slide).

taxa de compressão:

comprimento original da cadeia: s

comprimento comprimido da cadeia: l

t = (s-l)/s

restrições para codificação:

cada palavra de código deve corresponder a um único símbolo: deve haver unicidade na interpretação.

comprimento de uma palavra de código para símbolos mais frequentes deve ser menor que o para símbolos menos frequentes: redução do código

durante a interpretação do código comprimido, a interpretação deve ser sequencial: não há necessidade de retroceder na cadeia interpretada.

**Codificação de Huffman**

(no pdf)

Exercícios:

Araras comem banana nanica em Araraquara

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| .1 | Q1 | A13 | R5 | S1 | \_5 | C2 |
| O1 | M3 | B1 | E2 | N4 | I1 | U1 |

S1 - 00000

O1 - 00001

.1 - 00010

Q1 - 00011

E2 - 0010

C2 - 0011

N4 - 010

R5 - 011

B1 - 100000

I1 - 100001

U1 - 10000

M3 - 1011

\_5 - 100

A13 - 11

Leitura:

Drozdek (2002). Capítulo 11: Compressão de dados. Seções 11.1 e 11.2