**Ricardo Bernardelli – 405.2941-1**

**Compressão de dados**

A compressão de dados é o ato de reduzir o espaço ocupado por dados num determinado dispositivo. Essa operação é realizada através de diversos algoritmos de compressão, reduzindo a quantidade de Bytes para representar um dado, sendo esse dado uma imagem, um texto, ou um arquivo (ficheiro) qualquer.

Comprimir dados destina-se também a retirar a redundância, baseando-se que muitos dados contêm informações redundantes que podem ou precisam ser eliminadas de alguma forma. Essa forma é através de uma regra, chamada de código ou protocolo, que, quando seguida, elimina os bits redundantes de informações, de modo a diminuir seu tamanho nos ficheiros. Por exemplo, a sequência "AAAAAA" que ocupa 6 bytes, poderia ser representada pela sequência "6A", que ocupa 2 bytes, economizando 67% de espaço.

Além da eliminação da redundância, os dados são comprimidos pelos mais diversos motivos. Entre os mais conhecidos estão economizar espaço em dispositivos de armazenamento, como discos rígidos, ou ganhar desempenho (diminuir tempo) em transmissões.

Embora possam parecer sinônimos, compressão e compactação de dados são processos distintos. A compressão, como visto, reduz a quantidade de bits para representar algum dado, enquanto a compactação tem a função de unir dados que não estejam unidos. Um exemplo clássico de compactação de dados é a desfragmentação de discos.

**Classificações**

**Com perdas e sem perdas**

Esta é a forma mais conhecida de se classificar os métodos de compressão de dados. Diz-se que um método de compressão é sem perdas (em inglês, lossless) se os dados obtidos após a compressão são idênticos aos dados originais, ou os dados que se desejou comprimir. Esses métodos são úteis para dados que são obtidos diretamente por meios digitais, como textos, programas de computador, planilhas eletrônicas, etc., onde uma pequena perda de dados acarreta o não funcionamento ou torna os dados incompreensíveis. Um texto com letras trocadas, uma planilha com valores faltantes ou inexatos, ou um programa de computador com comandos inválidos são coisas que não desejamos e que podem causar transtornos. Algumas imagens e sons precisam ser reproduzidos de forma exata, como imagens e gravações para perícias, impressões digitais.

Por outro lado, algumas situações permitem que perdas de dados poucos significativos ocorram. Em geral quando digitalizamos informações que normalmente existem de forma analógica, como fotografias, sons e filmes, podemos considerar algumas perdas que não seriam percebidas pelo olho ou ouvido humano. Sons de frequências muito altas ou muito baixas que os humanos não ouvem, detalhes muito sutis como a diferença de cor entre duas folhas de uma árvore, movimentos muito rápidos que não conseguimos acompanhar num filme, todos estes detalhes podem ser omitidos sem que as pessoas percebam que eles não estão lá. Nesses casos, podemos comprimir os dados simplesmente por omitir tais detalhes. Assim, os dados obtidos após a compressão não são idênticos aos originais, pois "perderam" as informações irrelevantes, e dizemos então que é um método de compressão com perdas (em inglês, lossy).

**Simetria**

Quando falamos de métodos simétricos ou assimétricos de compressão de dados, estamos falando na verdade nas diferenças de complexidade entre a compressão e a descompressão. Quando a compressão e a descompressão são feitas executando-se métodos ou algoritmos idênticos ou bem semelhantes, dizemos que o método de compressão é simétrico. Bons exemplos são os algoritmos de codificação aritmética, ou o método LZW, baseado em dicionários. Neles o algoritmo de compressão e de descompressão são praticamente idênticos, e tem a mesma complexidade.

Quando o método de compressão é mais complexo que o de descompressão (ou em casos raros, o de descompressão é mais complexo que o de compressão), dizemos que o método de compressão é assimétrico. Este tipo de método é útil quando vamos comprimir apenas uma vez, mas descomprimir várias vezes (músicas em formato MP3 são um bom exemplo disso). Nesse caso temos todo o tempo do mundo para comprimir, mas a descompressão tem de ser feita em tempo real. Algoritmos como o LZ77 ou o DEFLATE, usados em programas comuns de compressão de dados são tipicamente assimétricos.

**Adaptabilidade**

A compressão de dados pode ser baseada em métodos rígidos, cujas regras não variam de acordo com os dados, nem a medida que os dados são lidos. São os métodos não adaptativos. Por outro lado, diversos métodos conseguem ir se adaptando aos dados a medida que estes são processados. Nesse caso o método é adaptativo. Métodos baseados em dicionário como as famílias LZ77 e LZ78 são naturalmente adaptativos já que é inviável que os programas de compressão carreguem dicionários de dados padronizados, ou que os dicionários sejam enviados junto com os arquivos. Métodos de aproximação de entropia, como a codificação de Huffman ou a codificação aritmética podem ser tanto adaptativos, inferindo as probabilidades a medida que os dados são lidos, como não adaptativos, usando probabilidades fixas ou determinadas por uma leitura prévia dos dados. Codificações fixas, como os códigos de Golomb, são naturalmente não adaptativos.

**De fluxo e de bloco**

Tradicionalmente os métodos de compressão de dados tratam os dados como um fluxo contínuo de dados, em geral visando à transmissão dos dados ou seu armazenamento seqüencial. Entretanto métodos mais recentes se aproveitam do fato de que dados próximos uns dos outros podem ser agrupados e processados em conjunto, de forma a aumentar a compressão. Esses métodos são classificados como métodos de compressão de bloco (block compression em inglês), como, por exemplo, o método de Burrows-Wheeler. Os métodos tradicionais são conhecidos como métodos de compressão de fluxo (stream compression em inglês).

**Classificação quanto à operação**

Os métodos de compressão são também classificados pela forma como operam, e pelo objectivo que buscam atingir. Nesta forma de classificação podemos citar:

* métodos estatísticos ou métodos de aproximação de entropia: São métodos que usam as probabilidades de ocorrência dos símbolos no fluxo de dados e alteram a representação de cada símbolo ou grupo de símbolos. Desta forma eles visam reduzir o número de bits usados para representar cada símbolo ou grupo de símbolos, e assim aproximar o tamanho médio dos símbolos ao valor da entropia dos dados. São os métodos que se baseiam diretamente na teoria da informação como a codificação de Huffman ou acodificação aritmética.
* métodos baseados em dicionários, ou métodos de redução de redundância: São os métodos que usam dicionários ou outras estruturas similares de forma a eliminar repetições de símbolos (frases) redundantes ou repetidas. Métodos como LZ77 e LZ78 fazem parte desse grupo. Os programas mais usados de compressão sem perdas em geral associam uma técnica baseada em dicionário com uma técnica estatística.
* Transformações são métodos que por si só não comprimem os dados, mas são capazes de transformar dados que não seriam comprimidos ou seriam pouco comprimidos pelos métodos normais, em dados que podem ser mais facilmente comprimidos. São em geral usados para compressão com perdas de forma a eliminar a correlação entre os dados adjacentes e assim identificar quais dados podem ser eliminados sem prejuízo do resultado final. Um exemplo é a Transformada discreta de cosseno, usado pelos padrões JPEG e MPEG. Uma exceção é o método de Burrows-Wheeler, usado na compressão sem perda de dados.