

COMPACTAÇÃO DE ARQUIVOS UTILIZANDO O ALGORITMO DE HUFFMAN E ÁRVORES BINÁRIAS

Gustavo Roncolato SOARES¹
Paula Piloto LANGHI²

RESUMO: O presente resumo tem por objetivo realizar uma abordagem sobre estruturas de dados relacionadas à codificação, bem como demonstrar, a partir da técnica de codificação de Huffman, conceitos relativos à compressão de dados através de uma árvore binária.

Palavras-chave: Compactação de arquivos. Algoritmo de Huffman. Árvores Binárias.

1 INTRODUÇÃO

A partir da evolução tecnológica e surgimento da internet, a necessidade de se obter informações em tempo hábil com gasto mínimo de espaço torna-se cada vez maior. Apesar disto, deve-se atentar para que não haja perda de qualidade dos dados a serem transmitidos..

Dessa forma, a compressão de dados é um recurso muito empregado e que permite aprimorar consideravelmente os dados a serem enviados, resultando em diminuição do tempo de transmissão, apesar da limitação de velocidade que a rede ainda apresenta.

2 DESENVOLVIMENTO

¹ Discente do 2º ano do curso de Sistemas de Informação do Centro Universitário “Antonio Eufrásio de Toledo” de Presidente Prudente. e-mail gustavoroncolato@hotmail.com.

² Docente do curso de Sistemas de Informação do Centro Universitário “Antonio Eufrásio de Toledo” de Presidente Prudente. Mestre em Ciências Cartográficas pela Universidade Estadual Paulista. e-mail paula_langhi@toledoprudente.edu.br. Orientador do trabalho.

Arquivos como fotos, vídeos, livros e músicas geram grande volume de dados e, por isso, tornam-se iminentes à aplicação de técnicas de compactação de dados que possibilitem reduzir a quantidade de informações necessárias, garantindo a exatidão dos dados e facilitando, conseqüentemente, seu armazenamento, transmissão e processamento

Estruturas de dados consistem em uma organização de dados na memória do computador de modo que possam ser processados de maneira eficiente. Para isso, os algoritmos e organizações de dados devem ser aplicados de forma coerente e racional, a fim de que haja otimização destes dados no intuito de solucionar problemas extremamente complexos e extensos.

Dentre as formas de organizações existentes, destaca-se a árvore, que, segundo Pinto (1990), “é uma representação capaz de exprimir a relação de hierarquia existente entre os dados, ou a relação de composição onde um conjunto de dados é subordinado a outro”.

Tal conceito pode ser mais facilmente compreendido utilizando-se como exemplo uma árvore genealógica, a qual expressa uma relação hierárquica entre membros de uma família e que, a partir de uma geração parental, permite a compreensão de informações das gerações subsequentes.

Tratando-se de árvores em computação, os seguintes itens devem ser determinados:

- NÓ, que é pai de todos os demais, chama-se RAIZ, e sempre estará no nível 0;
- Cada NÓ subsequente ao NÓ PAI irá formar outro subconjunto de zero ou mais elementos;
- Sempre quando o NÓ não possuir filhos, é denominado NÓ FOLHA, compreendendo-se, dessa forma, que este é o último elemento.

As árvores binárias são especiais, pois quando ordenadas, permitem a condução de pesquisa, inserção e exclusões de maneira eficaz.

Para a criação desse modelo de árvore, é necessário realizar as seguintes observações:

- Se a árvore for vazia, o símbolo é colocado na raiz;
- Se o elemento for menor que o símbolo da raiz, ele é colocado na subárvore da esquerda; se for maior, é colocado à direita.

Em meados da década de 50, David Huffman propôs um método muito utilizado nos dias atuais para compactação e descompactação de diversos formatos de arquivo, conhecido como Árvore de Huffman.

O método estatístico criado por Huffman permite a atribuição de um código binário aos arquivos a serem comprimidos. O comprimento de cada parte do código não é idêntico para todas as partes do arquivo a ser compactado, sendo que as partes mais utilizadas são codificadas de forma menor, enquanto as mais raras recebem códigos binários mais longos.

Esta codificação é baseada na “Teoria da Informação”, modelo de códigos sem perda de dados e caracterizado por uma sequência de símbolos genéricos. O código é totalmente reversível e permite descompactação, resultando em um fluxo de dados exatamente igual ao fluxo de origem.

Quando implementado, o código de Huffman resulta na construção de uma árvore binária “bottom-up”, cuja posição varia de acordo com a probabilidade de repetição. Os ramos da esquerda são codificados com 0 e os da direita com 1.

Quando se percorre da raiz da árvore até as suas folhas, obtém-se o código binário correspondente a uma parte do arquivo. No processo de decodificação, também se faz o uso dessa mesma árvore. No entanto, os passos realizados são o inverso da codificação.

3 CONCLUSÃO

Dessa forma, a partir da utilização do código de Huffman é possível analisar que o arquivo original possui 12,4 KB, e quando é aplicado o algoritmo para a codificação de compressão, o arquivo é reduzido para 9,75 KB, ocorrendo um ganho de quase 3 KB, tendo o arquivo texto em teste apenas uma linha. Nesse

sentido observa-se a eficácia do método para compressão de uma grande quantidade de dados sem que ocorra perda de dados originais.

REFERÊNCIAS BIBLIOGRÁFICAS

Algoritmo de Huffman para Compreensão de Dados. USP. Disponível em: <https://www.ime.usp.br/~pf/estruturas-de-dados/aulas/huffman.html>. Acesso em: 07 setembro. 2018.

Algoritmo de Huffman. UNIVERSIDADE FERNANDO PESSOA. Disponível em: <http://multimedia.ufp.pt/codecs/compressao-sem-perdas/codificacao-estatistica/algoritmo-de-huffman>. Acesso em: 07 setembro. 2018.

Compactação Sem Perda pelo Método de Huffman Associado à Transformada de Wavelet. Lucas Carlos Arrigoni. Disponível em: https://www.uvv.br/edital_doc/COMPACTA%C3%87%C3%83O%20SEM%20PERDA%20PELO%20M%C3%89TODO%20DE%20HUFFMAN%20ASSOCIADO%20%C3%80%20TRANSFORMADA%20DE%20WAVELET.pdf. Acesso em: 07 setembro. 2018

Estruturas de Dados. UNICAMP. Disponível em: <http://www.dca.fee.unicamp.br/cursos/EA876/apostila/HTML/node10.html>. Acesso em: 06 setembro. 2018.

O que são Estruturas de Dados. Boson Treinamentos. Disponível em: <http://www.bosontreinamentos.com.br/estruturas-de-dados/estruturas-de-dados>. Acesso em: 06 setembro. 2018.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília: Senado, 1988.

CENTRO UNIVERSITÁRIO “ANTONIO EUFRÁSIO DE TOLEDO” de Presidente Prudente. **Normalização de apresentação de monografias e trabalhos de conclusão de curso**. 2007 – Presidente Prudente, 2007, 110p.