

Huffman Coding and Decoding

Josemar Rocha da Silva
Professor: Herbert Rocha

Huffman Coding

— — —

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.

Carac	Freq	Cód
espaço	7	111
a	4	010
e	4	000
f	3	1101
h	2	1010

Origem

Foi a monografia de David A. Huffman onde foi desafiado por seus professores Robert M. Fano e Claude Shannon, criadores da codificação Shannon-Fano. Na monografia proposta, Huffman deveria encontrar a codificação binária mais eficiente.

Ele então criou uma codificação que faz uso de árvore binária utilizando frequências relativas.

Algoritmo

— — —

```
enquanto tamanho(alfabeto) > 1:
    S0 := retira_menor_probabilidade(alfabeto)
    S1 := retira_menor_probabilidade(alfabeto)
    X := novo_nó
    X.filho0 := S0
    X.filho1 := S1
    X.probabilidade := S0.probabilidade + S1.probabilidade
    insere(alfabeto, X)
fim enquanto

X = retira_menor_símbolo(alfabeto) # nesse ponto só existe um símbolo.

para cada folha em folhas(X):
    código[folha] := percorre_da_raiz_até_a_folha(folha)
fim para
```

Aplicação

Utilizada em aplicações de compressão como GZIP, PKZIP e até mesmo formatos de imagens como JPEG e PNG.

Huffman é simples, rápido e não possui patente.

Complexidade de Tempo

$O(n \log n)$, onde n é o número de caracteres únicos; Se existem n nodos, a função para extrair um nodo da heap mínima será chamada $2 \cdot (n-1)$ vezes, esta função leva um tempo de $O(\log n)$, pois ela chama uma outra função chamada Heapify. Assim, o tempo em geral, tem uma complexidade de $O(n \log n)$.