

Huffman-Verfahren

Prinzip:

Dem häufigsten Zeichen wird der kürzest mögliche Code, dem seltensten der optimal längste Code zugeordnet.

Zeichen, die nicht vorkommen, werden nicht verändert.

Algorithmus:

Die Häufigkeit jedes vorkommenden Zeichen wird ermittelt.

Ein binärer Baum, aus dem sich der Huffman-Code jedes einzelnen Zeichens ableiten lässt, wird erzeugt:

Jedes einzelne Zeichen stellt zunächst einen eigenen Baum dar.

Vereinige solange die beiden Bäume mit der geringsten Zeichenwahrscheinlichkeit bis nur noch ein einziger Baum übrig bleibt.

Wenn man den Baum von der Wurzel zu jedem einzelnen Zeichen verfolgt und dabei an bei Verzweigung in den linken Teilbaum eine 0 und bei einer Verzweigung in den rechten Teilbaum eine 1 zuordnet, ergibt sich ein optimaler und eindeutiger Code für jedes einzelne Zeichen.

In der Ursprungsdatei wird jedes Zeichen durch seinen neuen Code ersetzt.

Die Information, welcher Code welchem Zeichen entspricht wird an die Datei angehängt. Meist werden dort noch weitere diverse statistische und technische Parameter gespeichert. Dadurch vergrößert sich die Datei wieder geringfügig.

Anwendung:

- Textkompression
- Jpeg
- Mp3
- ...

siehe auch:

www.swisseduc.ch