

# Informatik

## Der Huffman-Algorithmus

David Huffman (1925-1999) entwickelte im Jahre 1952 ein Verfahren zur verlustlosen Kompression von Daten.

Verlustlos bedeutet, dass die Ursprungsdaten nach der Kompression originalgetreu wiederhergestellt werden können.

David Huffman (1925-1999) entwickelte im Jahre 1952 ein Verfahren zur verlustlosen Kompression von Daten.

Verlustlos bedeutet, dass die Ursprungsdaten nach der Kompression originalgetreu wiederhergestellt werden können.

Statt jedem Zeichen einen gleichlangen Code zuzuordnen (beispielweise 8 Bit im Falle von ASCII), bekommen Zeichen, die häufiger im Text vorkommen, einen kürzeren Code als seltene Zeichen.

Problematisch bei Codes mit unterschiedlichen Codewortlängen ist, dass im allgemeinen eine eindeutige Dekodierung von Zeichenfolgen nicht möglich ist.

# Aufgabe aus dem Informatik Biber 2009

## Codierung von Buchstaben

Biber codiert Buchstaben mit den zwei Ziffern "0" und "1" auf folgende Weise:

R = "1"  
S = "011"  
T = "010"

So steht zum Beispiel der Code "01011011" für die Zeichenkette "TRRS".  
Nun möchte Biber seinem System einen weiteren Buchstaben "U" hinzufügen.  
Er braucht dazu für "U" eine Codierung, die keine Mehrdeutigkeiten zulässt.  
Er kann dafür z. B. nicht "11" nehmen, weil sonst "RR" und "U" den gleichen Code "11" hätten.

**Auf welche Weise kann Biber den Buchstaben "U" eindeutig codieren?**

- A) mit U = "101"
- B) mit U = "110"
- C) mit U = "01110"
- D) mit U = "00"

# Aufgabe aus dem Informatik Biber 2009

## Codierung von Buchstaben

Biber codiert Buchstaben mit den zwei Ziffern "0" und "1" auf folgende Weise:

R = "1"  
S = "011"  
T = "010"

So steht zum Beispiel der Code "01011011" für die Zeichenkette "TRRS".  
Nun möchte Biber seinem System einen weiteren Buchstaben "U" hinzufügen.  
Er braucht dazu für "U" eine Codierung, die keine Mehrdeutigkeiten zulässt.  
Er kann dafür z. B. nicht "11" nehmen, weil sonst "RR" und "U" den gleichen Code "11" hätten.

Auf welche Weise kann Biber den Buchstaben "U" eindeutig codieren?

- A) mit U = "101"
- B) mit U = "110"
- C) mit U = "01110"
- D) mit U = "00"

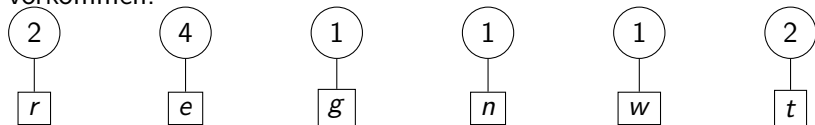
D) ist richtig

**Fano-Bedingung:** Wenn kein Codewort Anfangswort eines anderen Codewortes ist, dann ist jede codierte Zeichenreihe eindeutig dekodierbar.

Der Huffman-Algorithmus führt zu einer Codierung, die die Fano-Bedingung erfüllt und die beweisbar die größtmögliche Komprimierung erreicht.

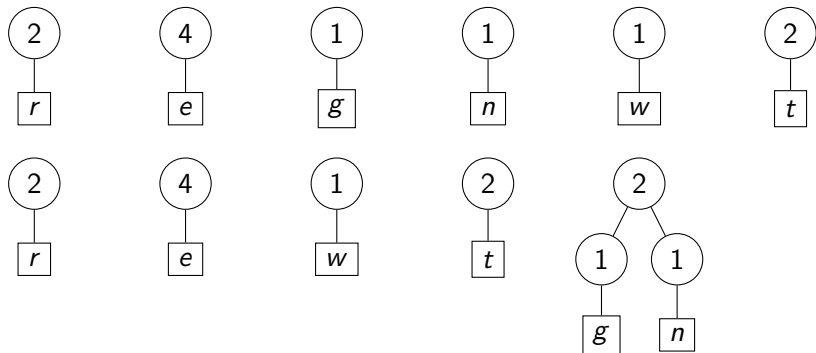
Wir wollen das Wort *regenwetter* mit dem Huffman-Algorithmus codieren.

Wir wollen das Wort *regenwetter* mit dem Huffman-Algorithmus codieren. Dazu zählen wir, wie oft die verschiedenen Buchstaben in dem Text vorkommen.

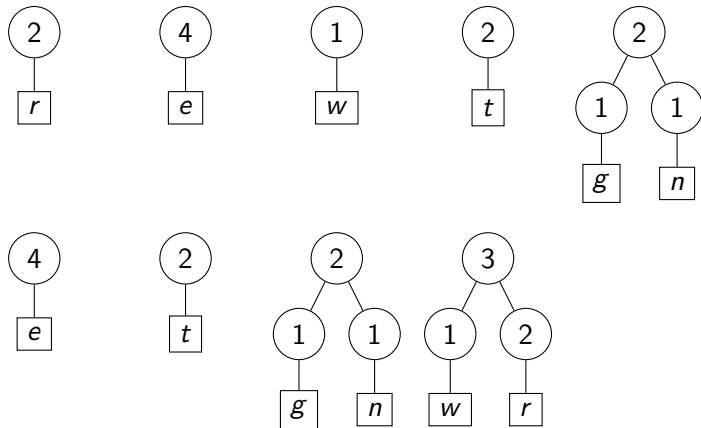


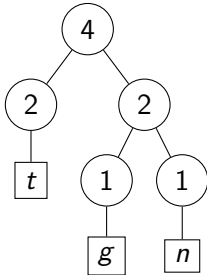
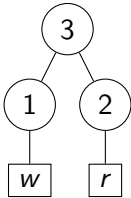
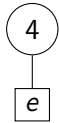
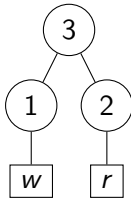
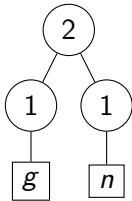
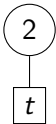
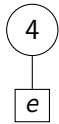
Wir betrachten das Ergebnis als eine Liste von Bäumen. Von links gehen wir durch die Liste und suchen die beiden Bäume mit den kleinsten Zahlen in der Wurzel. Das sind in unserem Fall die Bäume, an denen das g und das n hängen. Diese werden linker und rechter Teilbaum eines neuen Baums, in dessen Wurzel wir die Summe der beiden Zahlen schreiben. Der neue Baum wird rechts an die Liste angefügt. Die beiden benutzten Bäume verschwinden aus der Liste.

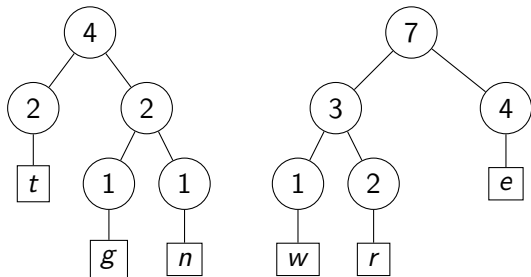
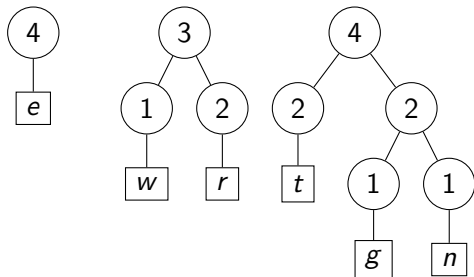


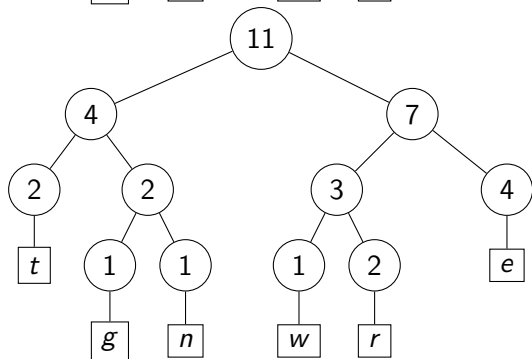
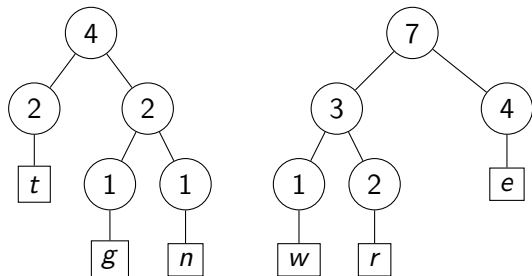


Wir wiederholen diesen Schritt, bis nur noch ein Baum übrigbleibt.

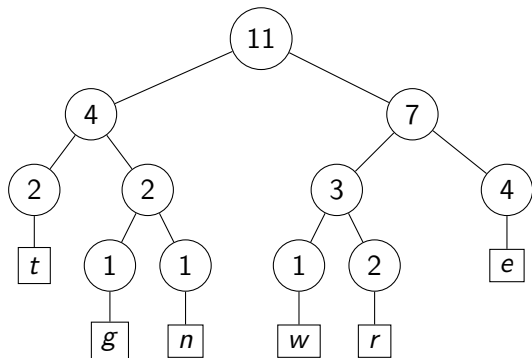








Wenn nur noch ein Baum vorhanden ist, können wir die Bitmuster für die Codierung der Zeichen ablesen. Abstieg in den linken Teilbaum ergibt 0, Abstieg in den rechten Teilbaum ergibt 1.



Wir erhalten folgende Codierung:

t = 00, g = 010, n = 011, w = 100, r = 101, e = 11

t = 00, g = 010, n = 011, w = 100, r = 101, e = 11

Die Codierung von *regenwetter* hat eine Länge von 27 bit:

101110101101110011000011101

Bei der ASCII-Codierung wird jedes Zeichen mit 8 bit codiert. Für die 11 Buchstaben von *regenwetter* wären also 88 bit erforderlich. In der Huffman Codierung werden 61 bit, das sind ca. 69% der ursprünglichen Codierung eingespart.

Zur Dekodierung muss natürlich der verwendete Huffman-Baum zur Verfügung stehen.