

Entropie-Codierung

Huffman

Codierung mit variabler Länge

Statistisches Verfahren:

Prinzip:

Häufig vorkommende Zeichen
durch kurze Bitfolgen darstellen

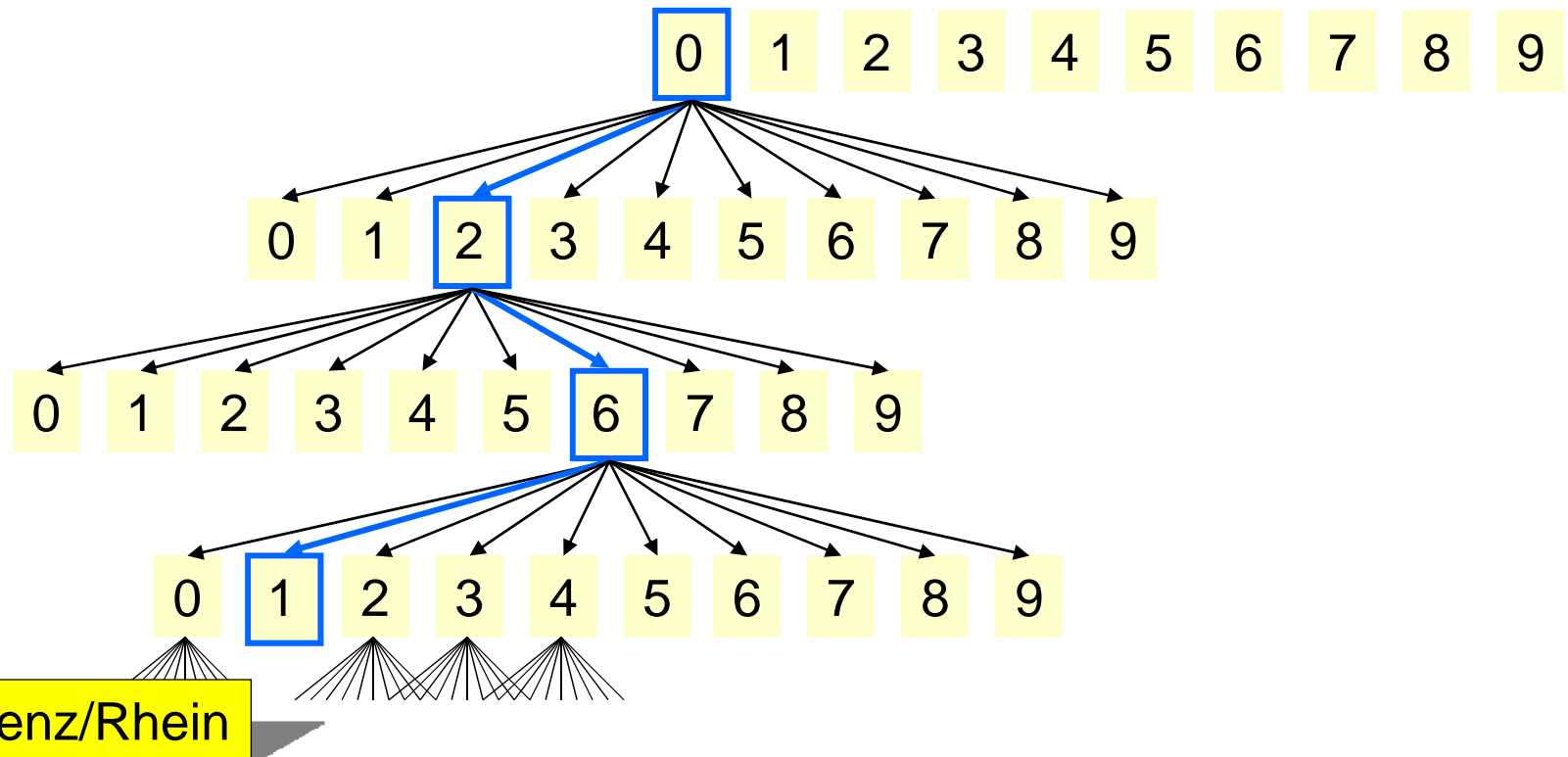
Präfix-Codes

Telefon-Nummern und Vorwahlen sind Präfix-Codes

0261 Koblenz/Rhein

Es gibt keine Vorwahlen 02610..., 02611..., 02612,...

Große Orte haben kurze Vorwahlen, kleine längere

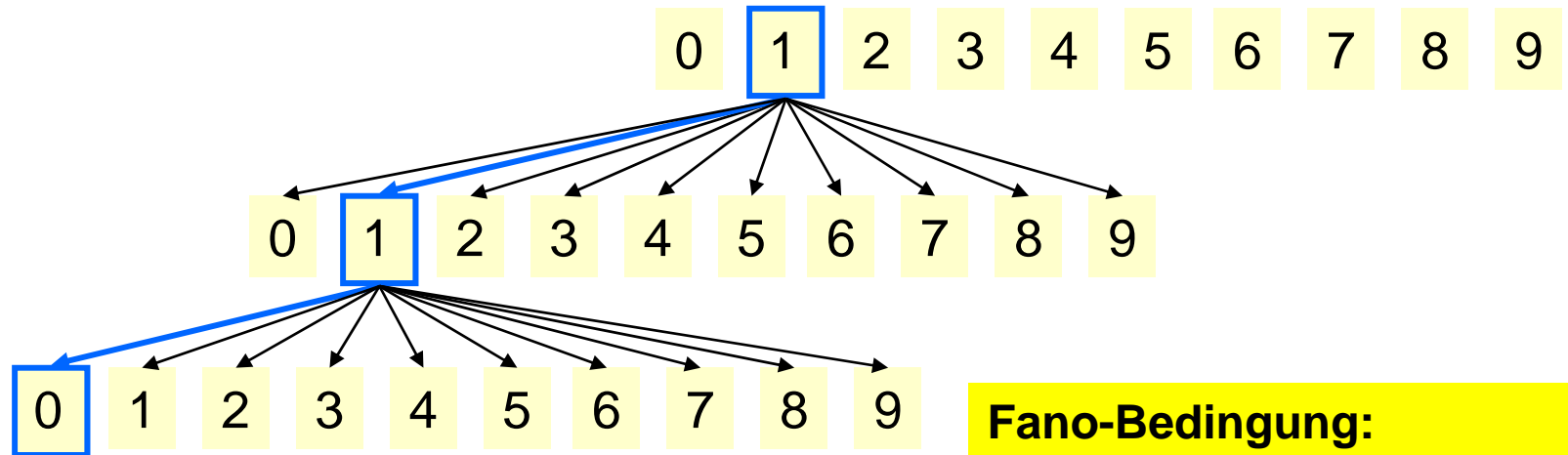


Präfix-Codes

Telefon-Nummern und Vorwahlen sind Präfix-Codes

110 Notruf

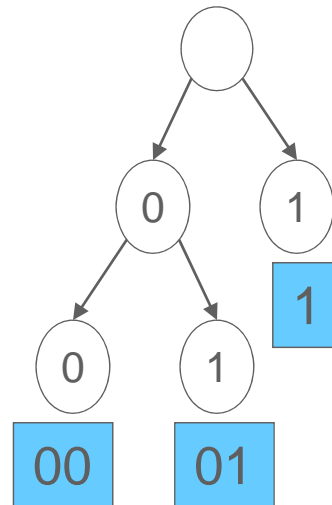
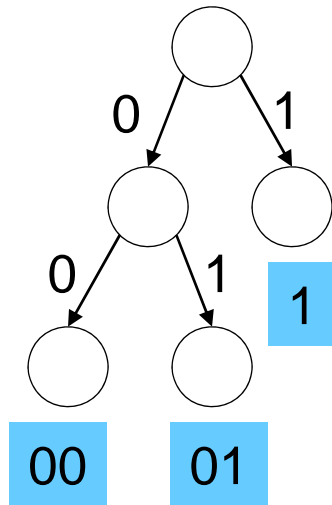
Es gibt keine Rufnummern 110...



Polizei

Fano-Bedingung:
Wenn kein Codewort Präfix eines anderen Codewortes ist, dann ist jede codierte Zeichenfolge decodierbar.

Binäre Präfixcodes



Oder so ...

Zeichenfolge

abacabaa

besteht aus

Code
s

5 a	1
2 b	00
1 c	01

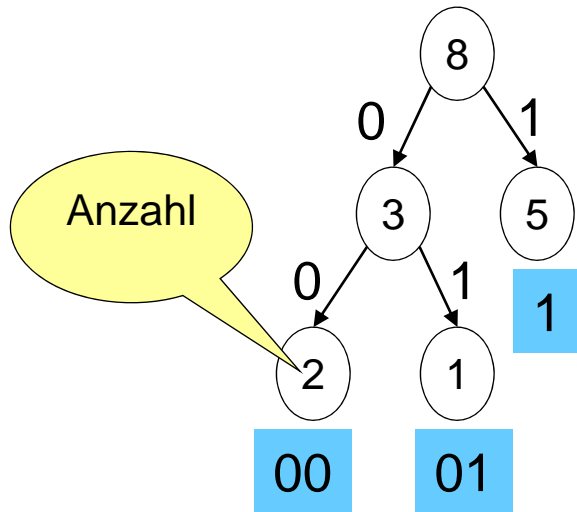
Konkatenation der
Kantenmarkierungen
von der Wurzel bis
in die Blätter
liefert Präfixcodes !

Codierung für **abacabaa**

10010110011

Ist das die kürzeste Präfixfolge ?

Binäre Prefixcodes



Konkatenation der
Kantenmarkierungen
von der Wurzel bis
in die Blätter
liefert Prefixcodes !

$$H = -\sum_{i=1}^m p_i \log_2 p_i$$

m Anzahl Zeichen
 p_i Wahrscheinlichkeit
des Auftretens des
Zeichens i

Zeichenfolge

abacabaa

besteht aus

$$H = -\left(\frac{5}{8} \log_2 \left(\frac{5}{8}\right) + \frac{2}{8} \log_2 \left(\frac{2}{8}\right) + \frac{1}{8} \log_2 \left(\frac{1}{8}\right)\right)$$

= 1,3..

Codierung für **abacabaa**

10010110011

Huffman-Algorithmus

Input: Zeichenfolge aus einem Alphabet

Output: Binärer Baum, Knotenmarkierung=Häufigkeiten,
Kantenmarkierungen 0,1

Start: Zähle die Häufigkeit jedes Zeichens im Text,
erzeuge für jedes Zeichen einen Knoten markiert mit der Häufigkeit

Wiederhole, bis es keinen Knoten gibt, zu dem keine Kante führt:

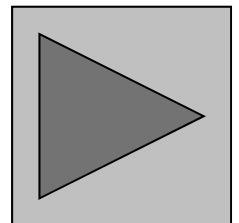
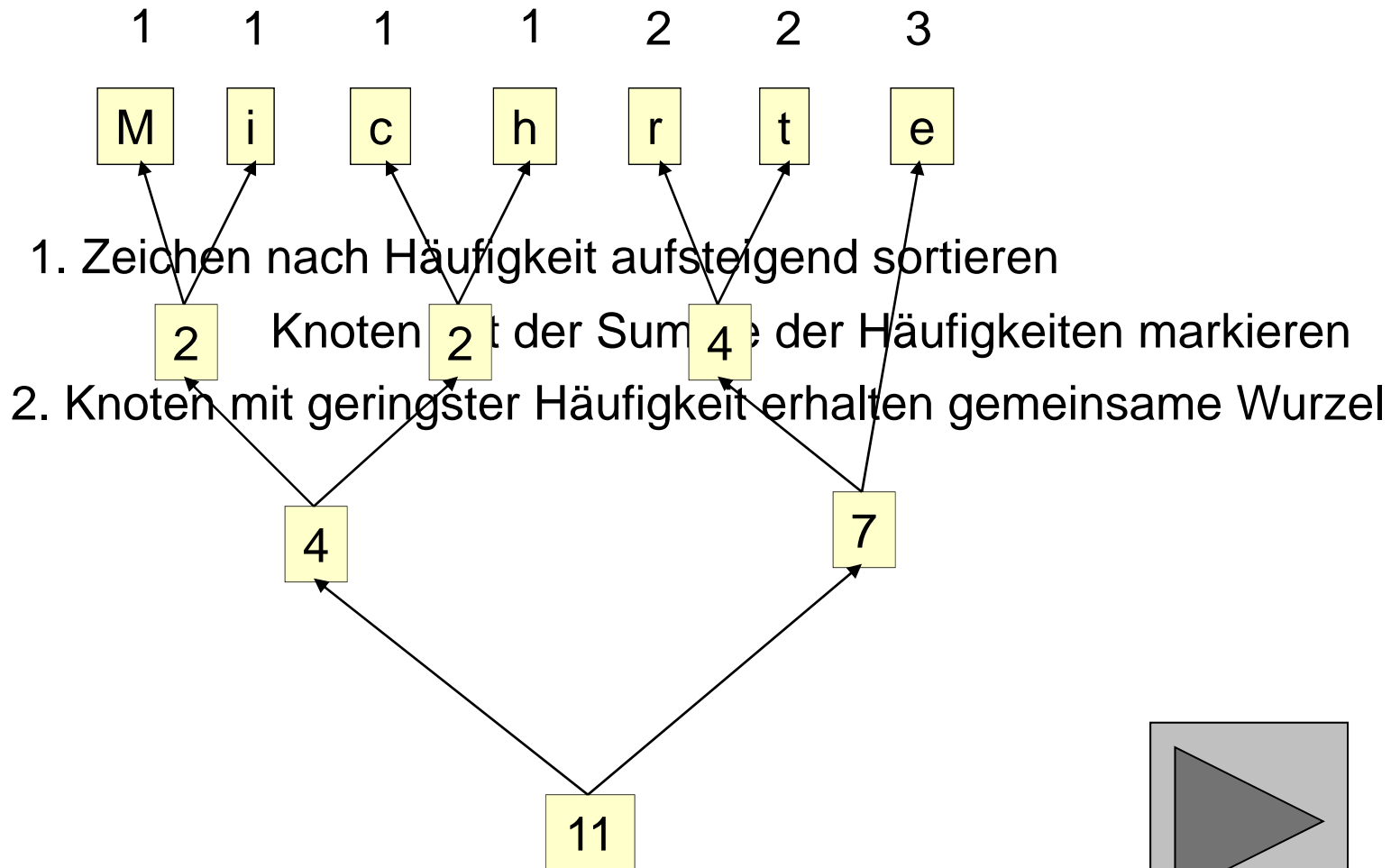
Wähle 2 Knoten, zu denen keine Kante hinführt, mit kleinster Häufigkeit.

Erzeuge neuen Knoten mit der Summe der Häufigkeiten der gewählten Knoten.

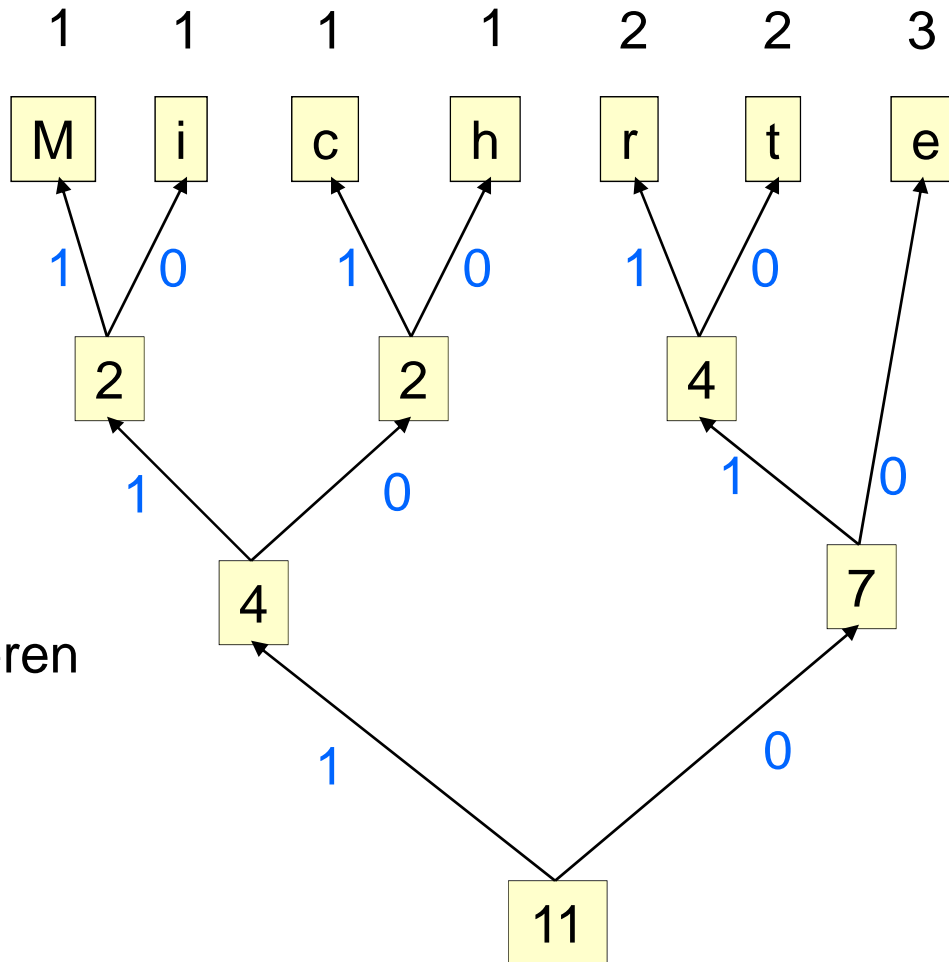
Ziehe zwei Kanten von dem neuen Knoten zu den gewählten.

Markiere die zwei von einem Knoten ausgehenden Kanten unterschiedlich mit 0 und 1.

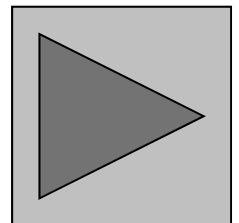
Huffman-Code „Meerrettich“



Huffman-Code „Meerrettich“

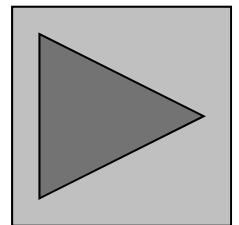
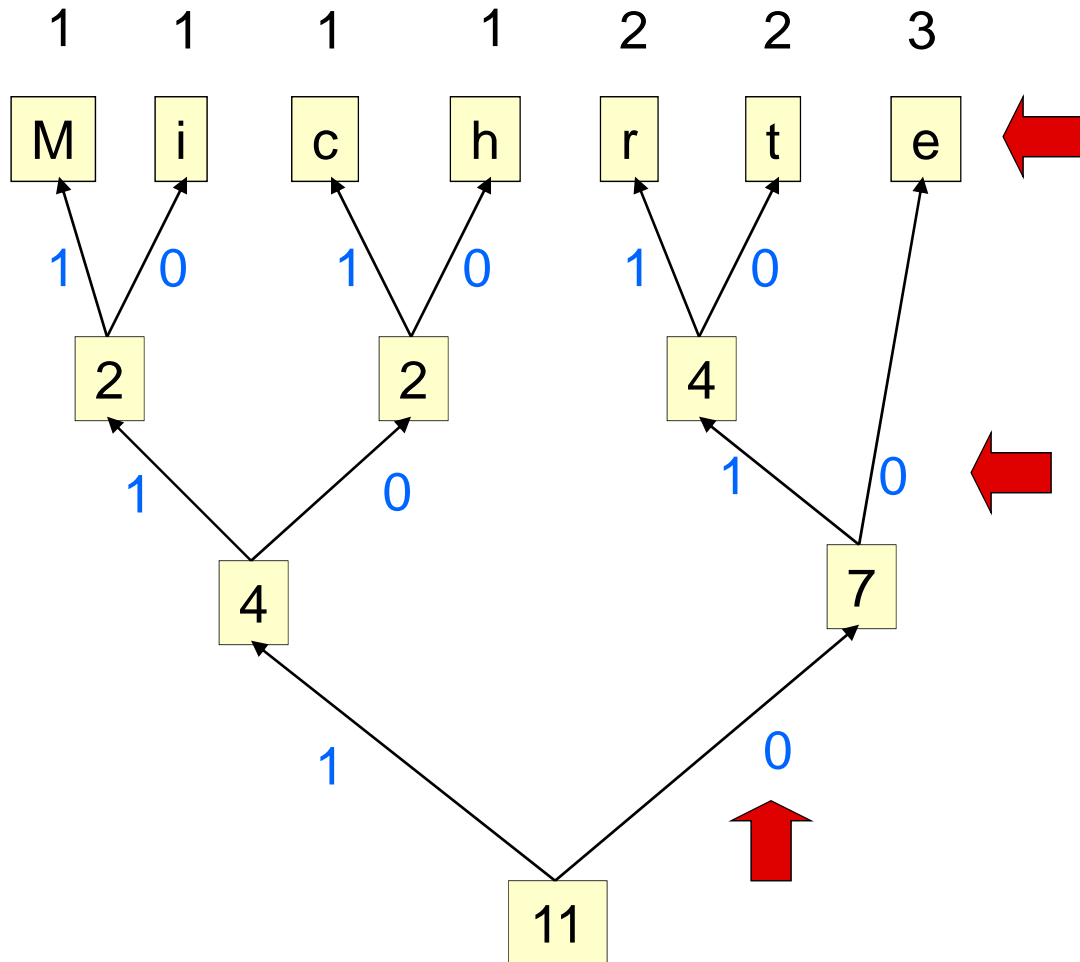


Kanten markieren

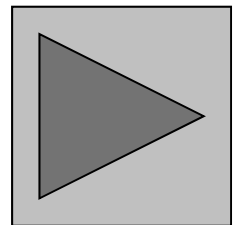
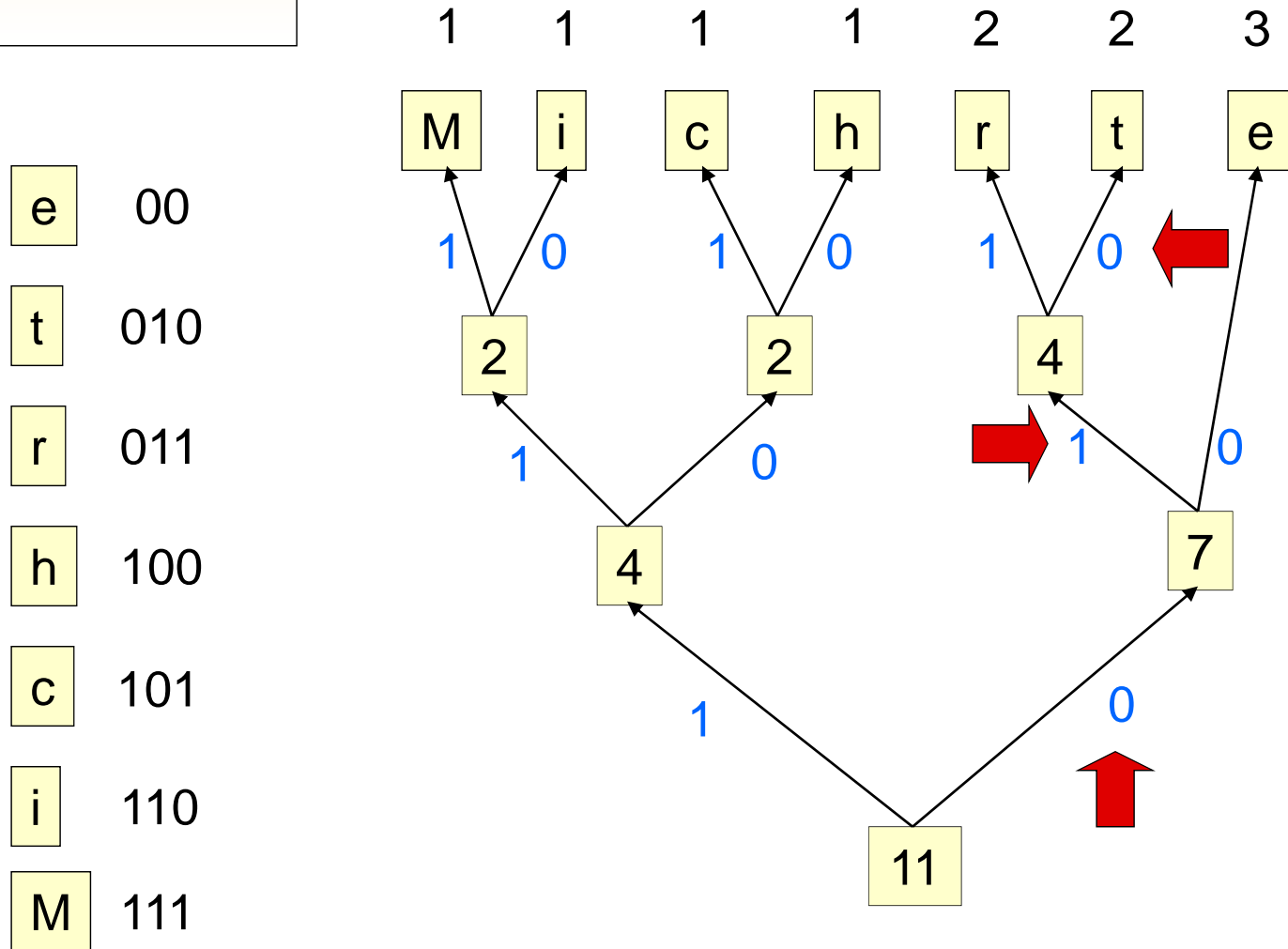


Huffman-Code „Meerrettich“

e 00
t
r
h
c
i
M



Huffman-Code „Meerrettich“



Meerrettich

111 00 00 011 011 00
010 010 110 101 100

Literatur zu diesem Kapitel:
Sedgewick: Algorithmen

Hyperlinks zu diesem Kapitel

Grafik-Quellen