

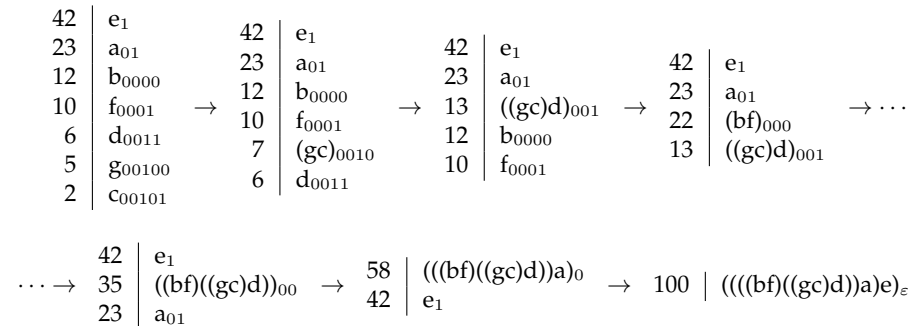
4. Studienbrief zur Informations- und Codierungstheorie

In dieser Woche schließen wir das Kapitel zur Quellencodierung ab, und zwar mit dem Huffman-Verfahren. Hierzu ein Beispiel.

Angenommen, im Quellalphabet treten sieben Zeichen verteilt auf wie durch die folgende Tabelle beschrieben:

Zeichen	a	b	c	d	e	f	g
Häufigkeit	23 %	12 %	2 %	6 %	42 %	10 %	5 %

Es soll eine Huffman-Codierung über dem Alphabet $A = \{0, 1\}$ hergestellt werden. Wir erwarten, daß e als häufigstes Symbol kurz codiert wird (und zum Beispiel b eher lang codiert wird). Dazu werden die Symbole nach Wahrscheinlichkeit absteigend sortiert und in jedem Schritt die beiden seltensten Symbole zu einem neuen, künstlichen Symbol zusammengefaßt mit neuer Gesamtwahrscheinlichkeit und neu geordnet, bis nur noch ein Symbol übrig bleibt. Auf die Notation der künstlichen Symbole kann man verzichten, doch muß dokumentiert sein, wie die einzelnen Zusammenfassungsschritte verlaufen sind. Man überlese zunächst die Subskripte am Ende der Symbole. Auch die künstlichen „Symbole“ (sehr lang notiert) muß man nicht „genau“ studieren.

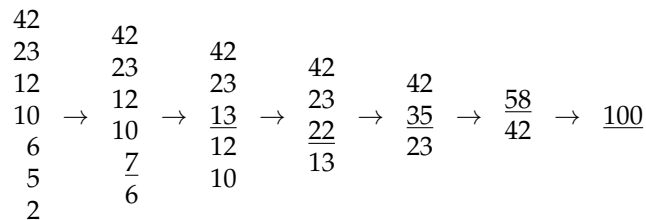


Am Ende steht bleibt eine Liste mit nur einem (künstlichen) Symbol, das mit 100 (Prozent) gewichtet ist (wenn sich niemand verrechnet). Dieses künstliche Wort symbolisiert die Wurzel des Codierungsbaums und kann künstlich mit ϵ codiert werden. Nun wird rückwärts verfahren: Solange es ein künstliches Symbol gibt, das schon künstlich durch ein Wort u codiert wurde, werden die beiden Symbole, aus denen es sich zusammensetzt (künstlich oder nicht) mit $u0$ und $u1$ codiert. Diesen Prozess kann man an den Subskripten

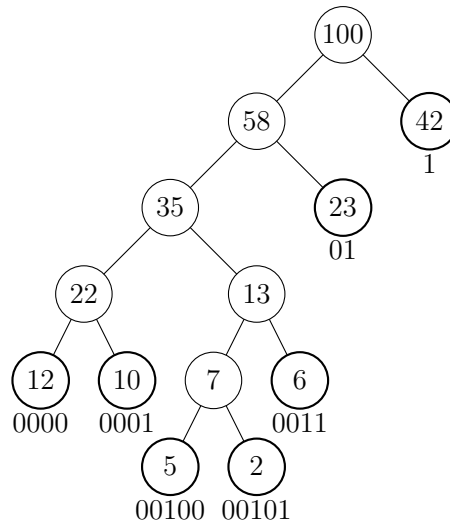
in der obigen Darstellung ablesen, wenn man die Schritte rückwärts liest. Die Huffman-Codierung ist dann die Einschränkung der künstlichen Codierung auf das Quellalphabet (das sind die Subskripte in der allerersten Tabelle):

Zeichen	a	b	c	d	e	f	g
Codewort	01	010	00101	0011	1	011	00100

Um das Verfahren von Hand durchzuführen, kann man einfach die Zusammenfassungsschritte durchführen und dabei den durch Zusammenfassung entstandenen Listeneintrag in der verkürzten Liste unterstreichen:



Hieraus kann man durch Rückwärtslesen einen beschrifteten Baum wie folgt gewinnen: Die Wurzel wird mit 100 beschriftet, die 100 bekommt zwei neue Nachbarn mit 58 und 42, danach die 58 zwei neue Nachbarn mit 35 und 23 etc. Allgemein bekommt das ausgezeichnete Element in der aktuellen Verteilung zwei neue Nachbarn, die mit den letzten zwei Einträgen der Vorgängerverteilung beschriftet sind. Diesen Baum zeichnet man gleich schichtweise, neue Nachbarn kommen links und rechts unter ihren Vorgänger.



Abschließend kann man die Codewörter an den Blättern durch den Links-Rechts-Verlauf von der Wurzel direkt ablesen und muß nur noch rekonstruieren, welche Buchstaben dadurch eigentlich codiert werden.

*

Lesepensum bis zum 15. Mai 2020:

Skript bis einschließlich Abschnitt 1.4, also das gesamte 1. Kapitel.

*

Übungpensum bis zum 15. Mai 2020:

Aufgaben 10 bis 18 (siehe Skript Abschnitt 1.5).

*

Fragen und Anregungen gerne per email an mich — IC am Anfang der Betreffzeile nicht vergessen!

Ilmenau, den 10. Mai 2020 · Matthias Kriesell