



## Modul C7 - Huffman-Code

### Zeitraumen

60 Minuten

### Zielgruppe

- Sekundarstufe I
- Sekundarstufe II

### Inhaltliche Voraussetzungen

C5 Codierung, Codebäume

### Lehrziel

Verständnis, dass Zeichen aus unserem Alphabet anderen Zeichen zugeordnet werden können (Grundprinzip der Codierung) sowie Optimierung von Codebäumen

### Requisiten

Mobilekarten [C\\_AB7.1](#), Schreibsachen

### Unterlagen

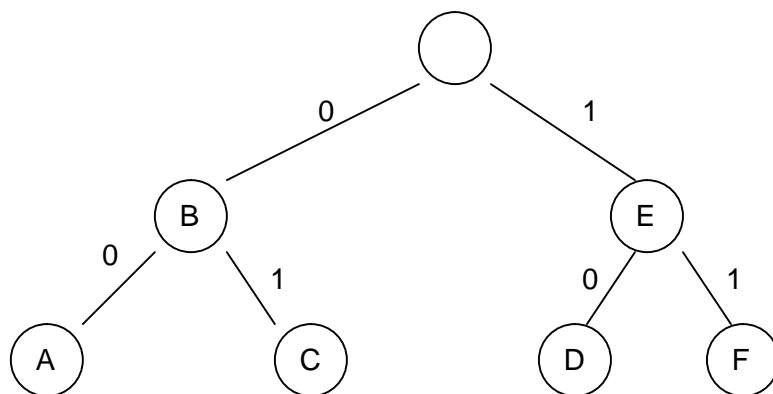
Arbeitsblatt [C\\_AB7.2](#)

### Partizipanden

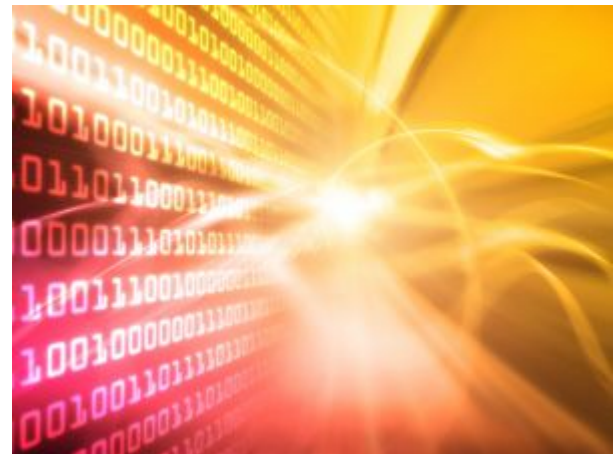
Gesamte Klasse

### Vorgehensweise

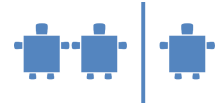
1. Besprechen oder Wiederholen Sie einige Grundlagen des Morsecodes (siehe C1, C5). **Wie könnte man den folgenden Code optimieren? Denkt dabei auch an den Morsecode...**



A	0
B	00
C	01
D	10
E	1
F	11



<http://www.sxc.hu/>



Sehen wir uns einmal den Codebaum an, wenn man auf die Häufigkeiten Rücksicht nimmt, mit denen die Buchstaben auftreten.

A	5,30
B	1,54
C	2,49
D	4,14
E	14,17
F	1,35

Welche Buchstaben sollten demnach die kürzesten Codes haben? - Sehr häufig sind E, gefolgt von A und D.

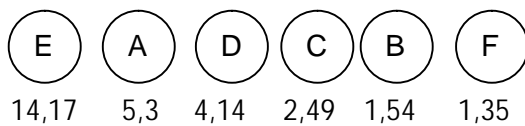
Welche dürfen den längsten Code haben? – B und F

Erstellen wir also einen neuen Codebaum, in dem E der Wurzel näher ist, als andere Buchstaben, beispielsweise B und F. Dabei erzeugen wir einen Code *variabler Länge*.

Wir gehen so vor, dass wir jeweils die Buchstaben mit der **geringsten Häufigkeit** zusammenfassen. Diese Buchstaben werden also in unserem fertigen Baum ganz unten zu finden sein. Daher wächst unser Baum von unten nach oben, bis wir dann das E ganz oben irgendwo dazu hängen können.

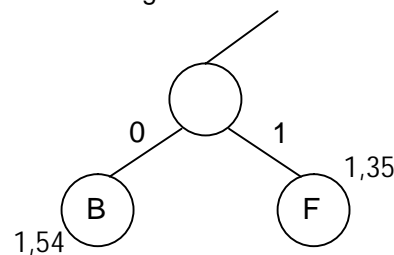
Weisen Sie die Schülerinnen und Schüler darauf hin, dass unser Baum von unten nach oben wächst. (siehe AB7.2)

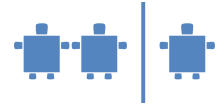
Es werden folgende Blätter mit den jeweiligen Buchstaben und Häufigkeiten entstehen:



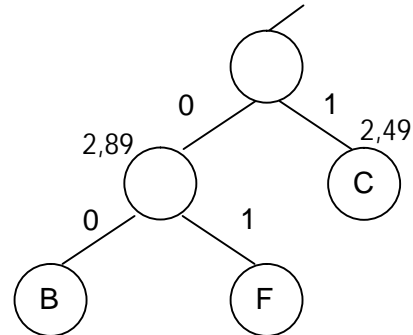
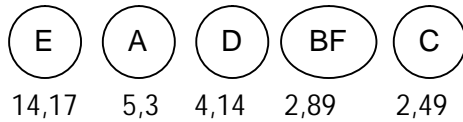
- a. Die Buchstaben mit der geringsten Häufigkeit wären im ersten Schritt B und F. Die Häufigkeit beträgt für den Teilbaum BF nun  $1,54+1,35 = 2,89$ . Der Teilbaum BF wird ab jetzt im Bezug auf die Häufigkeit wie ein einzelner Buchstabe behandelt.

A	5,30
BF	2,89
C	2,49
D	4,14
E	14,17



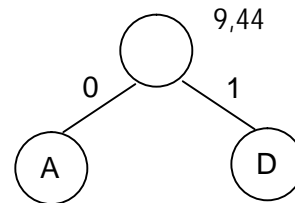


b. Die nächsten Buchstaben bzw. Teilbäume mit der geringsten Häufigkeit sind BF und C. Dabei gehen wir von den Werten der Tabelle in Punkt a. aus. Somit werden diese beiden zusammengefasst. Siehe unten:



A	5,30
BFC	5,38
D	4,14
E	14,17

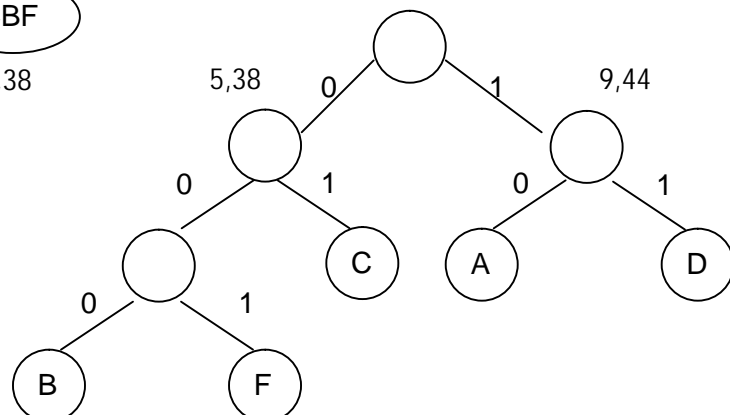
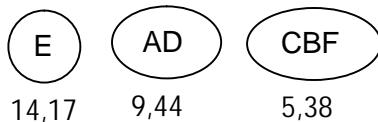
c. Wir gehen wieder zur Tabelle aus dem vorherigen Schritt b. und A und D haben die geringsten Häufigkeiten und diese beiden werden somit in einem ersten Schritt zusammengefasst. Sie ergeben einen neuen Teilbaum, den wir im nächsten Schritt erst an den bestehenden Teilbaum angliedern müssen.  
Neuer Ausgangspunkt:

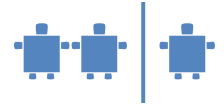


Ergebnis:

AD	9,44
BFC	5,38
E	14,17

d. Die beiden Teilbäume, die bis jetzt entstanden sind werden zusammengefasst, da die Teilbäume AD und BFC diejenigen mit der geringsten Häufigkeit sind wodurch nun diese beiden zusammengefasst werden.  
Neuer Ausgangspunkt:

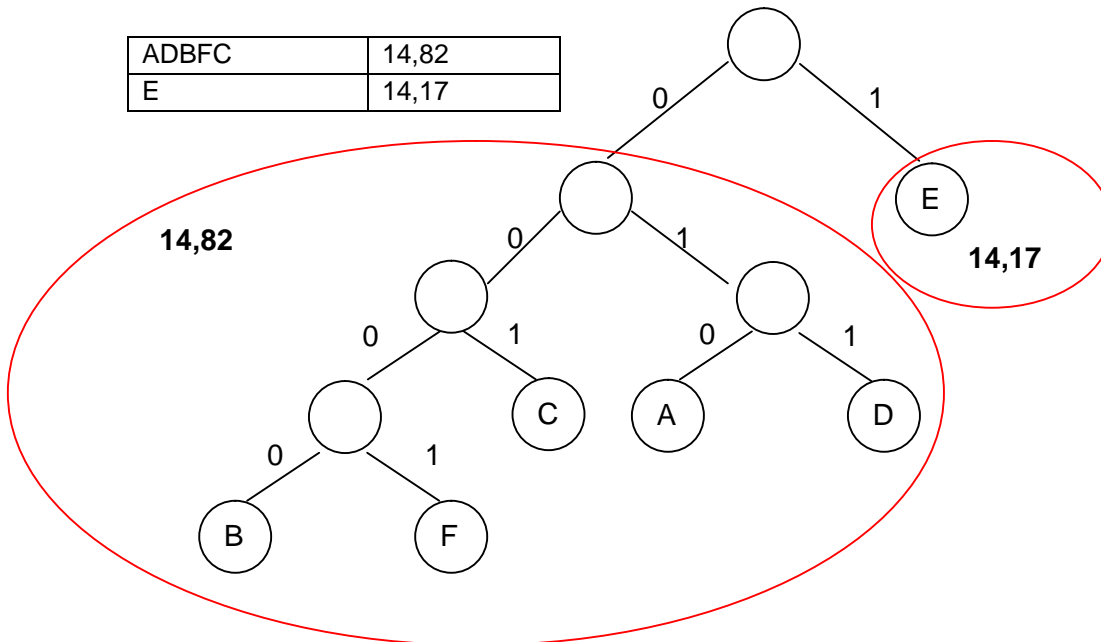




e. Übrig bleibt somit noch E.



ADBFC	14,82
E	14,17



Würde man sich den Baum als Mobile vorstellen und die dazugehörige Häufigkeit der Teilbäume als Gewichte, so kann man feststellen, dass der linke und der rechte Teilbaum in etwa ausgeglichen sind. Das Mobile würde also gerade hängen. [à Vorlage Mobile-Teile C\\_AB7.1.](#)

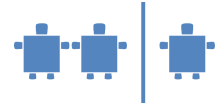
Man sieht, dass die Buchstaben mit hoher Häufigkeit bei diesem Verfahren im Baum weiter oben sind, als jene mit niedriger Häufigkeit.

Man nennt dieses Verfahren **Huffman-Codierung**, nach David Albert Huffman. Wir haben einen optimalen Code variabler Länge erzeugt.

Eine Bedingung die für die Eindeutigkeit eines Codes von Bedeutung ist, ist die **Fano-Bedingung**: Sie besagt, dass kein Codewort der Anfang eines anderen Codewortes sein darf.

Trifft dies für unseren Code zu? - ja

A	010
B	0000
C	001
D	011
E	1
F	0001



Wo würde das Problem liegen, wenn das nicht zuträfe? Nehmen wir als Beispiel die folgende Codierung:

A	101
E	01
I	100
O	11
U	11100

Die folgende Übertragungsfolge 10111100100 könnte bedeuten:

- 101 11100 100 a AUU
- 101 11 100 100 a AOII

Der Code ist also nicht eindeutig decodierbar!

Dieses Problem würde beim Morsecode auch auftreten.

Zusätzlich zu den beiden Zeichen kurz oder lang ist beim Morsecode noch ein Trennzeichen nötig.

Machen Sie die TN darauf aufmerksam indem Sie ein Beispiel verwenden:  
 z.B. AEIOU ( . \_ | . | .. | \_ \_ \_ | .. ) vs. ENEWDT ( . | \_ . | . | . \_ \_ | \_ .. | \_ )  
 Schreiben Sie die Folge von Signalen zuerst ohne Trennzeichen auf und lassen Sie die TN selbst die Interpretationsmöglichkeiten entdecken. Bei der Übertragung ist das Trennzeichen allerdings kein dritten Zeichen, sondern eine etwas längere Pause bis zum nächsten Signal.

**Quellen/Weiterführende Literatur**

Gallenbacher, Jens: Abenteuer Informatik. IT zum Anfassen von Routenplaner bis Online-Banking. Spektrum Akademischer Verlag, München, 2007.

Bollin, Andreas: Skriptum Einführung in die Informatik, Institut für Informatik-Systeme, Universität Klagenfurt.

Gumm, Heinz-Peter; Sommer, Manfred: Einführung in die Informatik. Oldenbourg Wissenschaftsverlag, München, 2002.