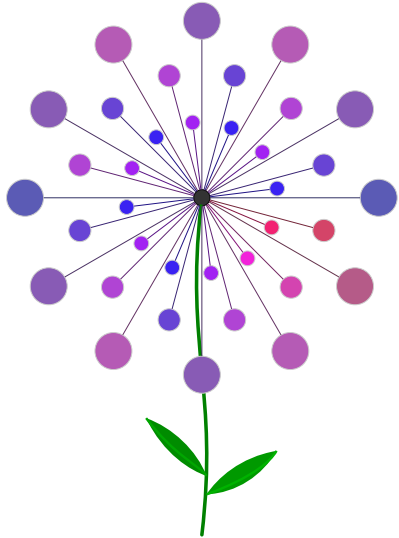


FreeFlower

Informatik: Data Science und Sicherheit

Kompression: Huffman-Kompression

Huffman-Kodierung

Idee: Immer zuerst Knoten mit kleinsten Prozentzahlen verbinden

Huffman-Kodierung: Algorithmus

- Jeder Buchstabe wird als Graph mit einem Knoten (=Baum mit einem Blatt) betrachtet. Die Häufigkeit jedes Buchstabens wird neben dem Blatt notiert. Wir wollen nun die Buchstaben zu Teilbäumen verbinden, bis wir einen einzigen Baum haben.

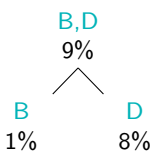


Huffman-Kodierung: Algorithmus

- Jeder Buchstabe wird als Graph mit einem Knoten (=Baum mit einem Blatt) betrachtet. Die Häufigkeit jedes Buchstabens wird neben dem Blatt notiert. Wir wollen nun die Buchstaben zu Teilbäumen verbinden, bis wir einen einzigen Baum haben.



- In einer Schleife: Wir schauen alle isolierten Blätter und Wurzeln von Teilbäumen an. Diejenigen Knoten, die die kleinste (gemeinsame) Prozentzahl enthalten werden miteinander verbunden. Die neue Wurzel wird markiert mit dem Prozentsatz beider Knoten.

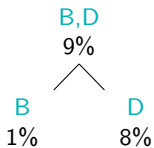


Huffman-Kodierung: Algorithmus

- Jeder Buchstabe wird als Graph mit einem Knoten (=Baum mit einem Blatt) betrachtet. Die Häufigkeit jedes Buchstabens wird neben dem Blatt notiert. Wir wollen nun die Buchstaben zu Teilbäumen verbinden, bis wir einen einzigen Baum haben.



- In einer Schleife: Wir schauen alle isolierten Blätter und Wurzeln von Teilbäumen an. Diejenigen Knoten, die die kleinste (gemeinsame) Prozentzahl enthalten werden miteinander verbunden. Die neue Wurzel wird markiert mit dem Prozentsatz beider Knoten.



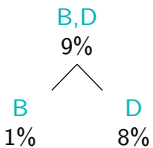
- Falls mehrere Möglichkeiten mit gleicher relativer Häufigkeit existieren, nehmen wir den Teilbaum mit der geringsten Tiefe.

Huffman-Kodierung: Algorithmus

- Jeder Buchstabe wird als Graph mit einem Knoten (=Baum mit einem Blatt) betrachtet. Die Häufigkeit jedes Buchstabens wird neben dem Blatt notiert. Wir wollen nun die Buchstaben zu Teilbäumen verbinden, bis wir einen einzigen Baum haben.



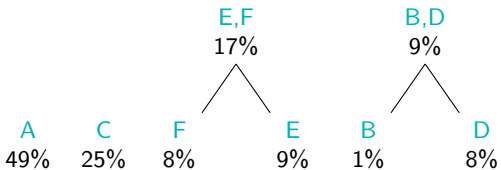
- In einer Schleife: Wir schauen alle isolierten Blätter und Wurzeln von Teilbäumen an. Diejenigen Knoten, die die kleinste (gemeinsame) Prozentzahl enthalten werden miteinander verbunden. Die neue Wurzel wird markiert mit dem Prozentsatz beider Knoten.



- Falls mehrere Möglichkeiten mit gleicher relativer Häufigkeit existieren, nehmen wir den Teilbaum mit der geringsten Tiefe.

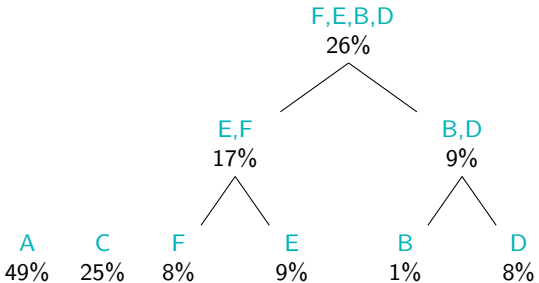
Welche Knoten werden als nächstes miteinander verbunden?

Huffman-Kodierung: Beispiel

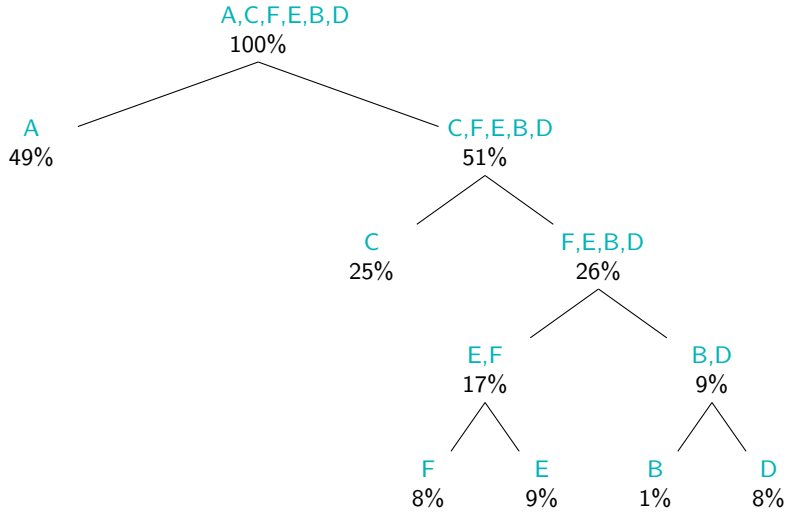


$\rightarrow$ Weshalb haben wir einen Baum $\{E,F\}$ statt $\{F,B,D\}$ gemacht?

Huffman-Kodierung: Beispiel

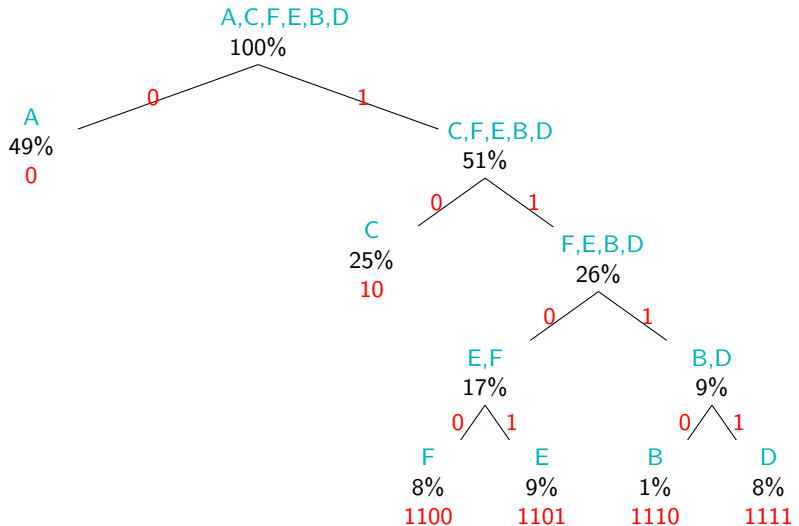


Huffman-Kodierung: Beispiel



→ Welcher Binärcode hat B? Welcher hat E?

Huffman-Kodierung: Beispiel



Fazit

- Strategie der **Maximale Balance**: *top-down*-Ansatz (von Wurzel zu Blättern), kann zu suboptimalen Kodierungen führen
- **Huffman-Kodierung**: *bottom-up*-Ansatz (von Blättern zur Wurzel), garantiert bestmögliche Kodierungslänge für alle Buchstaben

Auftrag

- Aufgaben 1.6 - 1.9