Huffman Coding

...

Prefix-free code & lossless data compression

Compression

Lossless

- Less effective
- Universal & domain Specific
- 100 % Quality remains

Examples:

- ZIP
- Huffman Coding
- Lempel-Ziv-Welch (LZW)
- PNG

Lossy

- More effective compression
- Domain specific only
- Quality lost

Examples:

- MP3
- JPEG
- H.264
- Dolby Digital

Text encoding: ASCII (*American Standard Code for Information Interchange*)

- 1 Character = 1 Byte
- Contains 127 different characters
- Does not contain Å Ä Ö

Example:

A = 0x*41* = 0b*0100 0001*

Fox = 0x*46 6F 78* = 0b*0100 0110 01100 01111 0111 1000*

Dec	*Hex*	*Char*
..	..	..
74	4A	J
75	4B	K
76	4C	L
..	...	..

Compression & Decompression

Huffman Compression

Original Data

Protocol

Compressed Data

“Voodoo”

Char	*Representation*
V	01
o	1
d	00

0b01110011

Prefix-free code

OK!

Char	Representation
V	01
o	1
d	00

Voodoo == 0b01110011

WRONG!

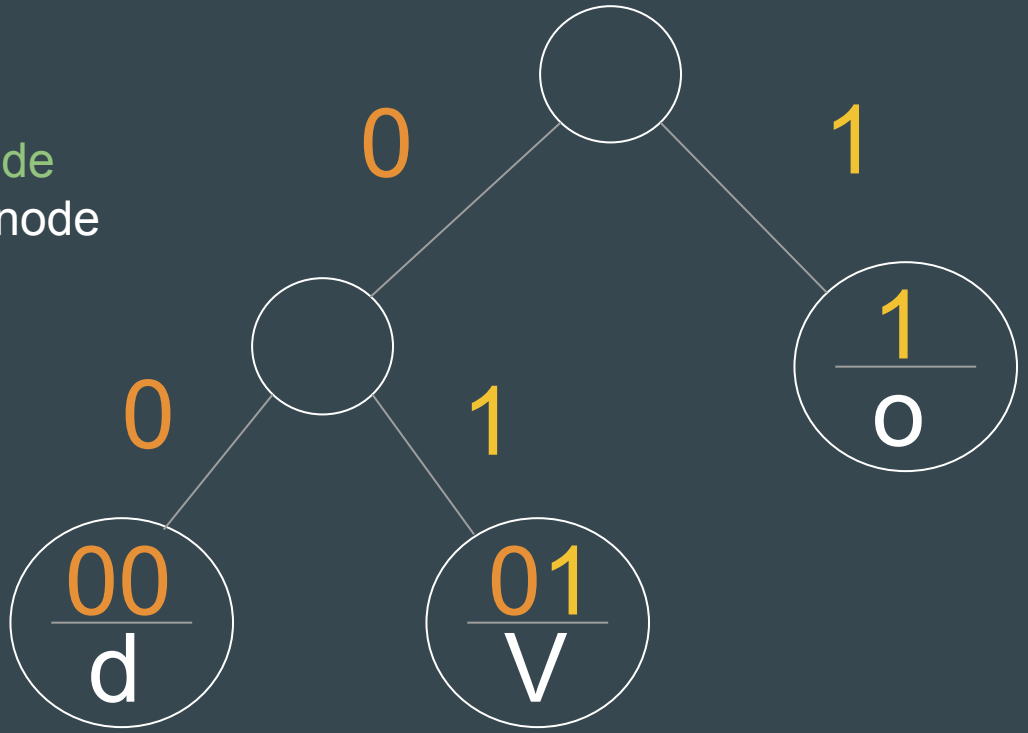
Char	Representation
V	11
o	1
d	01

Voodoo == 0b11110111
oVodoo == 0b11110111
oVodoV == 0b11110111

Huffman Tree

- Every leaf node is a character node
- A character node must be a leaf node

- Left == 0
- Right == 1



Huffman Tree

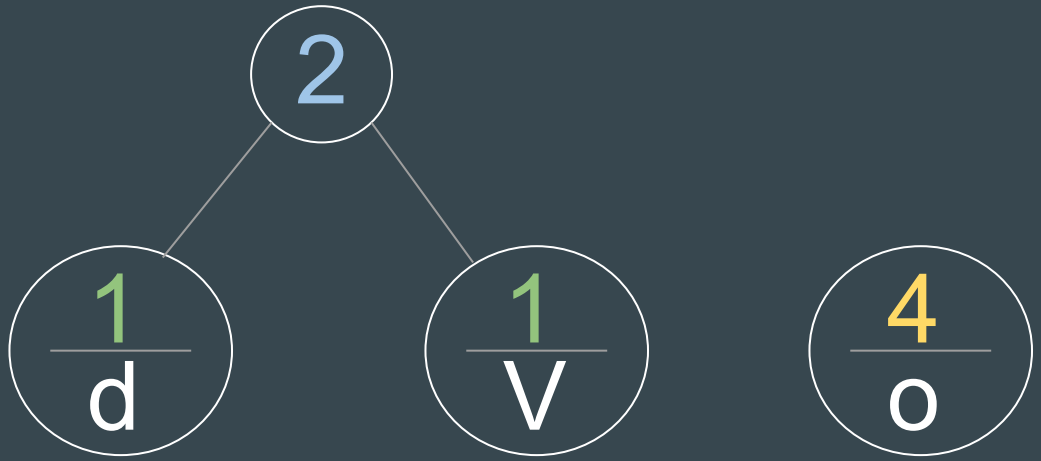
“Voodoo”

Char	*Occurrences*
V	1
o	4
d	1

Huffman Tree

“Voodoo”

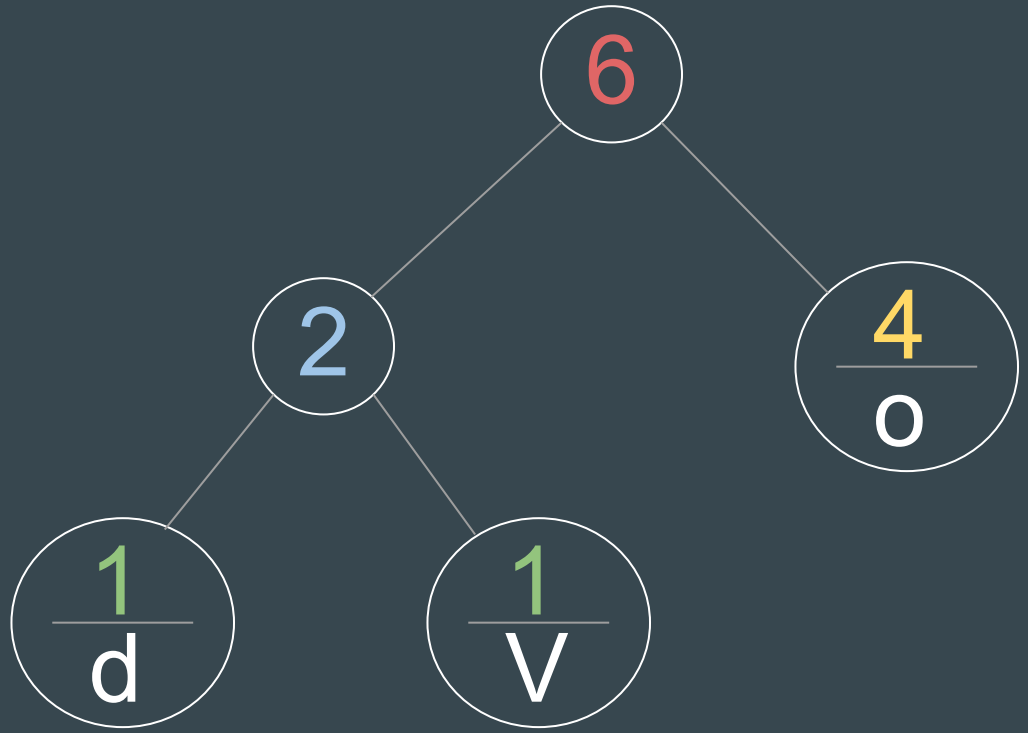
Char	*Occurrences*
V	1
o	4
d	1



Huffman Tree

“Voodoo”

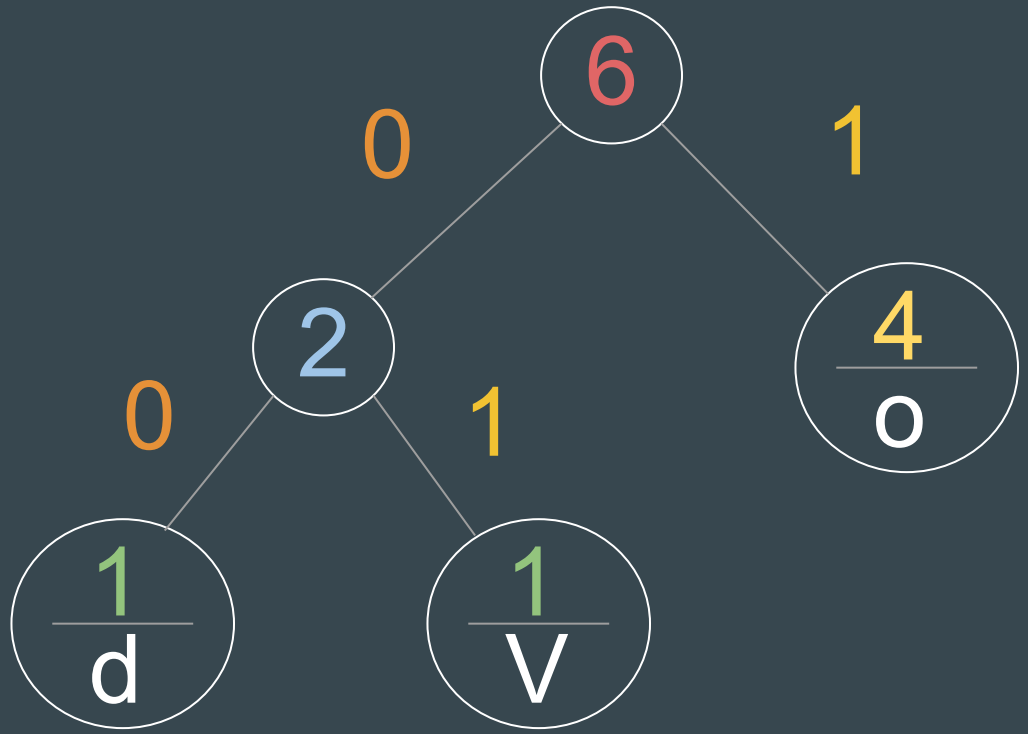
Char	*Occurrences*
V	1
o	4
d	1



Huffman Tree

“Voodoo”

Char	*Representation*
V	01
o	1
d	00



https://people.ok.ubc.ca/ylucet/DS/Huffman.html

Huffman Compression

1. Calculate Occurrences
2. Huffman Tree
3. Protocol
4. Compress