

DATA COMPRESSION:

Naive compression

Example: Text = ABRAKADABRA ($|Text| = 11$)

Latin1 coding: 1 char : 8 bits COMPRESSION:

CODE TABLE

Chars	Code
A	000
B	001
R	010
K	011
D	100

11 chars : 88 bits

1 char: 3 bits

11 chars: 33 bits

char by char
fixed code length

Text = A B R A K A ...

COMPRESSED Text = 000001010000011000...

COMPRESSED FILE: {COMPRESSED Text}

$$\Sigma = \{g_1, \dots, g_d\}$$

$$k = \lceil \log d \rceil \text{ bits in 1 char}$$

CODE TABLE

k bits for a single char. $2^k \geq 5 > 2^{k-1}$

$$k \geq \lceil \log 5 \rceil \rightarrow k = \lceil \log 5 \rceil = 3$$

MUFFMAN CODING

PREFIX-FREE CODE

ABRAKA
DABRA

Code table

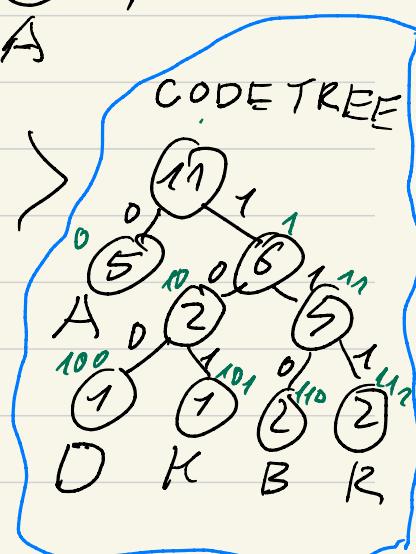
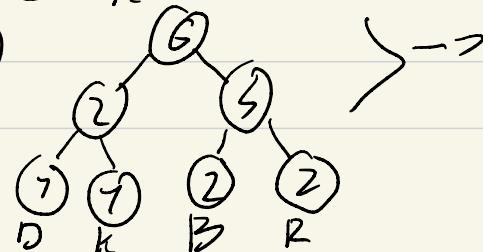
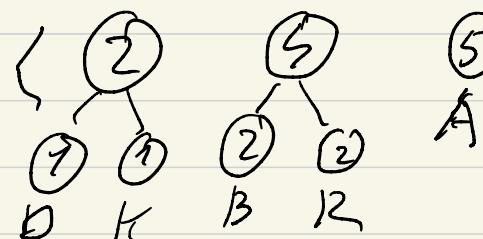
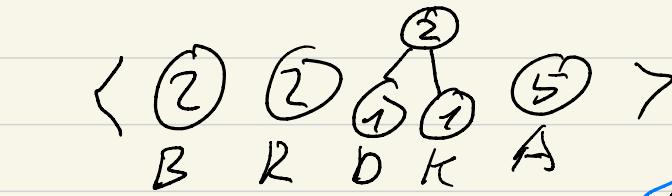
chars	occurrences	codes	
A	HHH = 5	0	5
B	II = 2	110	6
R	II = 2	111	6
K	I = 1	101	3
D	II = 1	100	3

|Compressed code| = 23 bits

Text = ABRAKA
DABRA

(char by char, varying code length)
frequent chars: shorter codes
rare chars: longer codes

minPrQ: < ①, ①, ②, ②, ⑤ >
D K B R A



COMPRESSED FILE

CODE TREE

DECOMPRESSED text

LOSSLESS COMPRESSION:

- Huffman compression
- LZW compression

LESS COMPRESSION:

Nano compression

HUFFMAN code

LZW compression

IMPLEMENTED

zip, unzip progs.

LÉMPEL-ZIV-WELCH (LZW) COMPRESSION

Example: $\Sigma = \{A, B, C\}$ | fixed code | $E[1..15]$ (compressed)

$T = A \underline{B} A \underline{B} A B A A C A A C C B B \underline{A} \underline{A} \underline{A} \underline{A} \underline{A} \underline{A} \underline{A} \underline{A}$

Dict. entry

Initial Dictionary (D)

out code	input word	next char	new code	letter	code
1	A	B	4	A	1
2	B	A	5	B	2
4	AB	A	6	C	3
6	ABA	A	7		
1	A	C	8		
3	C	A	9		
1	A	A	10		
8	AC	C	11		
3	C	B	12		
2	B	B	13		

out code	in. word	next char	new code
5	BA	A	14
10	AA	A	15
15	AAA	A	16
15	AAA	—	—

D is not included in the compressed file.

COMPRESSED CODE = {1, 2, 3, 6, 1, 3, 1, 8, 3, 2, 5, 10, 15, 15}

ALPHABET: $\Sigma = \{A, B, C\}$

in. code	out. word	next char	new code
1	A	B	?
2	B	A	5
3	AB	A	6
6	ABA	A	7
1	A	C	8
3	G	A	9
1	A	A	10
8	AC	C	11
3	C	B	12
2	B	B	13
5	BA	A	14

Initial D

letter	code
A	1
B	2
C	3

D entry
(sc : code)

in. code	out. word	next char	new code
10	AA	A	15
15	AAA	A	16
15	AAA	—	—

(C=?) : code (sc) = h & D

T = A B A B A B A A C A A C C B B A A A A A A A A
 $C = S_1$ uncompresssed text