

Slovníkové metody komprese dat

- slovník - $D = (M, C)$
- ↑
konečný set stringů (frází)
- ↑
mapování $M \rightarrow$ code words

LZ77 - sliding window method

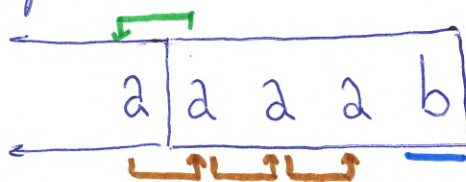
- search buffer (thousands of bytes) |S|
- look-ahead buffer (tens of bytes) |L|

$\rightarrow (i, j, a)$
 \uparrow offset \uparrow length \uparrow next symbol
 $\lceil \log_2 S \rceil$ $\lceil \log_2 L \rceil$
 - výrazně rychlejší dekomprese než komprese

[LZH - Huffman coding]

- offset - o kolik se vrátit zpět
- length - kolik potom vypsal symbolů
 - může být i více než 1
 - pokračují tím, co jsem sám vypsal
- další symbol

příklad



(1, 3, b)

varianty:

LZSS

- search buffer is BST, look-ahead buffer circular queue
- dokud nejsou dvě položky nebo nekomprimovaný symbol
 - pro každých 8 jímek byle s flagy

- 2 byte break even point (pokud nejde komprimovat, output raw ...)

- Deflate
- huffman coding
 - rozdelení kódu na $(i, j), (a)$
 - pro každý znaků stream

- LZMA:
- podobný deflate
 - variable dictionary size
 - range encoding

LZO - decompression speeds

LZH - worse ratios but extremely fast

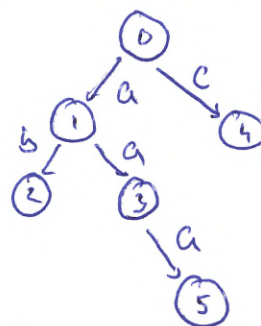
LZ78 - growing dict (trie = prefixový stream)

$\mapsto (i, a)$
 index \uparrow \uparrow next symbol

- ~~index~~ ^{index} kódování β kódem

$$\begin{bmatrix} \beta(0) = 0 & \beta(1) = 1 \\ \beta(2i+j) = \beta(i)\beta(j) \end{bmatrix}$$

aabaaacaaa



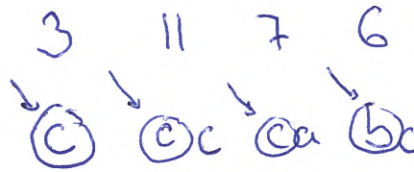
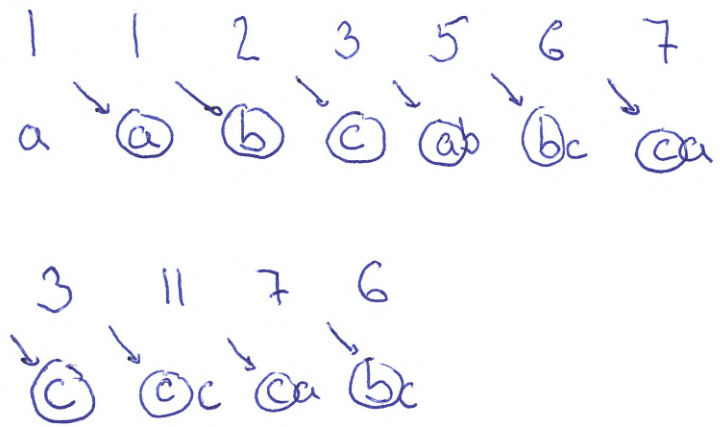
$(0, a)(1, b)(1, a)(0, c)(3, a)$

LZW

- (i) - pointer to dictionary
- všechny proměnné inicializované od začátku
 - poslední symbol freze bude první symbol další freze

increment string by:	add string to dict	
	per phrase	per input char
One char	LZW	LZH
several chars	LZMW	LZAP

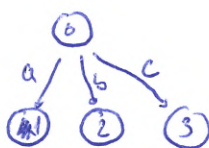
index	phrase	coded
0	ϵ	ϵ
1	a	a
2	b	b
3	c	c
4	aa	1a
5	ab	1b
6	bc	2c
7	ca	3a
8	abb	5b
9	bcc	6c
10	cac	7c
11	cc	3? \rightarrow 3c
12	ccc	11c
13	cab	7b



poslední symbol musí být
stejný jako ten první

proč \checkmark

funkce:



- vstup index, který má být vyhledán
- na něj připojíme další symbol ve vyhledání
- ten ale nepředstavuje! je prvním pro další slovo