Kontextové metody komprese dat

PPM — Prediction by partial matching

- trie for all found contexts
- each level for one context length

PPMA : escape symbol - pravděpodobnost $\frac{1}{1+M}$

PPMB : seeing twice is believing

PPMC : escape handled as others

PPM*

- dva různé kontexty, podle kterých se dá kódovat
- deterministický (vždy pokračuji jen jeden symbol)
  - → hledám nejkratší takový
  - → pokud není, použiji nedeterministický (jak u PPM)
- deterministický kontext může být neomezený (délka)

DCA

- zakázaná slova, antidictionaries
  - → minimální zakázané slovo

    $x[1...n-1]$ a $x[2...n]$ nejsou zakázaná slova
- komprese & dekomprese
  - → output pouze to, co se nedá jasně rozhodnout z AD
- konstrukce AD → strom → suffix linky do min. zakázaných slov
- synchronizace ?

verze

$\swarrow$ static

dynamic — jednoběhové, bez synchronizace, exceptions

almost antiword — gain je použit, ne všechna slova kódována
- 2 pass

ACB — Associate Coder of Buyanovsky

- unbounded context and content
  → pairs inserted into dictionary & sorted right to left (by the context)

- output (d, l, a) ⎯ first unmatched in lookahead buffer
  ⎯ number of matched symbols

distance between best context and content

1	swiss␣|m	context is swiss␣m
2	swi|ss␣m	→ best match for it is between 2 & 3
3	s|wiss␣m	(where would it appear sorted)
4	swis|s␣m	content is iss␣ i3␣ missing
5	swiss|␣m	→ best match is 4 symbols long at position 6
6	sw|iss␣m	

$$\rightarrow (6-2, 4, i)$$

další varianta: — context matching at least k symbols

$$\begin{array}{l} 4 \quad xx...x\ 0\ zz...z\ 0\ .... \\ 5 \quad xx...x\ 0\ zz...z\ 1\ .... \\ 5 \quad xx...x\ 1 \end{array}$$
(with b above first 0, $\neg b$ above second 0)

$$(4, b, l) \quad \text{length} (\neg b - b)$$
position, mismatching bit

Symbol ranking
— next symbol is often easy to guess (3 or 4 guesses)
— context C of size n for symbol R — position of R in the list