LEMPEL-ZIV-WELCH COMPRESSION

LZW-block decomposition

For $w \in \Sigma^+$, the *LZW-block decomposition of* w is the sequence $B_0, \ldots, B_{m-1}$ of subwords such that

(i) T_0 is the trie with root 0 and branches to nodes $1, \ldots, q$ labeled $a_1, \ldots, a_q$, where $\Sigma = \{a_1, \ldots, a_q\}$.

(ii) If trie T_n and subwords $B_0, \ldots, B_{n-1}$ are constructed such that $w = B_0 \cdots B_{n-1} v$ for some $v \in \Sigma^+$, then

$$B_n := \begin{cases} \text{longest non-empty prefix of } v \\ \text{labeling a path in } T_n \end{cases}$$

If $v \neq B_n$, then

$$T_{n+1} := \begin{cases} T_n \text{ extended by a branch from the} \\ \text{node reached by } B_n \text{ to a new node } |T_n|, \\ \text{labeled with the first letter of } v \text{ after } B_n. \end{cases}$$

LZW-compression (pointers only!)

If $w = B_0 \cdots B_{m-1}$ is the *LZW* block decomposition, then

$$LZW(w) = p_0 \cdots p_{m-1}$$

where p_i is the node reached in T_{m-1} by the path labeled B_i.

LZW-Decompression

(i) T_0 is known from Σ;

(ii) if T_n and $B_0 \cdots B_{n-1} p_n \cdots p_{m-1}$ are reconstructed:

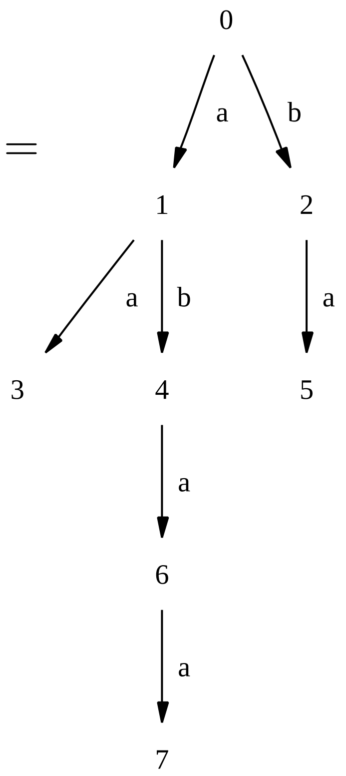
(iii) get B_n by reading the path from 0 to p_n in T_n;

(iv) get T_{n+1} by adding an edge from p_n to new node $|T_n|$, labeled by the first letter of B_{n+1}, which is the first label on the path to node $p_n \in T_n$, if $p_{n+1} \notin T_n$, else to p_{n+1}.

Expl. 12 LZW-compression of $w = aababababaaa$:

	B_0	B_1	B_2	B_3	B_4	B_5
$w =$	a	a	b	ab	aba	aa
$LZW(w) =$	1	1	2	4	6	3

$T =$



$$T_n = T \restriction \{k \mid k \leq n+2\}$$

Expl. 13 LZW-compression of $w = abbaabbabbb$:

	B_0	B_1	B_2	B_3	B_4	B_5	B_6	B_7
$w =$	a	b	b	a	ab	ba	bb	b
$LZW(w) =$	1	2	2	1	3	5	4	2