LZW ALGORITHM

TWO ALPHABETS:

* INPUT ALPHABET } # OF ELEMENTS
* "TABLE" ALPHABET } IS TYPICALLY A POWER OF 2

INPUT ALPH. C "TABLE" ALPH.

Ex. IN LECTURE NOTES: INPUT ALPH. = \{ ALL BYTES \}

256 SYMBOLS (0..255)

"TABLE" ALPH. = FIRST $2^{12}$
NON-NEGATIVE INTEGERS

INPUT ALPH. STREAM → ENCODER → COMPRESSED "TABLE" ALPH STREAM → DECODER → INPUT ALPH. STREAM

MUST BE THE SAME (LOSSLESS COMPRESSION)

LZW ENCODER:

"STATE":

(a) "TABLE" ("STRING TABLE")
A PARTIALLY DEFINED MAPPING
OF "TABLE" ALPH. TO WORDS IN THE
INPUT ALPH.

ALWAYS: INPUT ALPH. → CORRESPONDING
SINGLE SYMBOL WORD

(b) "STRING": EITHER "NONE" "EMPTY" OR AN EXISTING TABLE ENTRY (EMPTY)
INITIALIZED AT "NONE"

*UPDATE OF STATE/OUTPUT FOR THE NEXT INCOMING SYMB.*

IF STR + NEW IN TABLE:
    STR = STR + NEW
ELSE:
    SEND (SYMBOL (STR))
    ADD TO TABLE (STR + NEW)
Ex. 1  INPUT ALPH. = \{0, 1\}

TABLE ALPH. = \{0, 1, 2, 3, 4, 5, 6, 7\}

INPUT STREAM = "000 000 000 00"

<table>
<thead>
<tr>
<th>INPUT</th>
<th>S</th>
<th>TBL. INQ.</th>
<th>OUT</th>
<th>TABLE UPD.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>00</td>
<td>0</td>
<td>00</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>00</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0000</td>
<td>2</td>
<td>0000</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>00</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0000</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>0000</td>
<td>3</td>
<td>00000</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>00</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0000</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>0000</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>00000</td>
<td>4</td>
<td>000000</td>
</tr>
<tr>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

"COMPRESSED" STREAM:
"02340"

(WE ACTUALLY HAVE NOT WON ANYTHING YET)

LZW DECODER

STATE: (a) CURRENT TABLE (INITIALIZED IN ITS "MINIMAL" STATE)

(b) LAST SYMBOL TO ARRIVE (INITIALIZED AT "NOTHING")

UPDATE OF STATE/OUTPUT:

OUTPUT: TABLE (LAST SYMBOL)

WHEN "NEW", "LAST" ARE NOT "NOTHING"

ADD TO TABLE: TABLE (LAST SYMB.) + FIRST SYMB (TABLE (NEW))

LAST = NEW
Ex. 2

INPUT ALPH. = \{0, 1, ... 255\} \text{ (bytes)}

TABLE ALPH. = \{0, 1, ... 4095\} \text{ (typical)}

COMPRESSED STREAM = (5, 7, 7, 5, 253, 5)

ORIGINAL STREAM = ?

<table>
<thead>
<tr>
<th>IN</th>
<th>LAST</th>
<th>OUT</th>
<th>TABLE UPD.</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>5</td>
<td>256 : (5, 7)</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>7</td>
<td>257 : (7, 7)</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>7</td>
<td>258 : (7, 5)</td>
</tr>
<tr>
<td>253</td>
<td>5</td>
<td>5</td>
<td>259 : (5, 5)</td>
</tr>
<tr>
<td>5</td>
<td>259</td>
<td>(5, 5)</td>
<td>260 : (5, 5, 5)</td>
</tr>
</tbody>
</table>

RECONSTRUCTED STREAM = (5, 7, 7, 5, 5, 5, 5)

Ex. 3

IN THE SETUP OF EX. 2, IS IT POSSIBLE TO HAVE A COMPRESSION RATE OF 5000?

ANSWER: NO

WHY? IF THE TABLE ENCODES A SEQUENCE OF LENGTH \( n > 1 \), IT ALSO ENCODES A SEQUENCE OF LENGTH \( n-1 \)!

HENCE MAX. LENGTH OF A SEQUENCE FROM A TABLE IS 3841, IN WHICH CASE 12 BITS ENCODE 8.3841 BITS.

HENCE COMPRESSION RATE \leq \frac{8.3841}{12} \ll 5000