LZW ALGORITHM

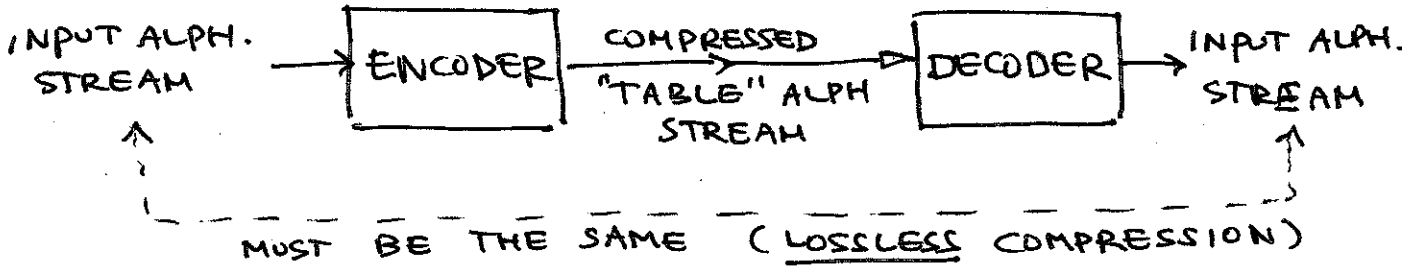
6.02 / Dec. 2, 2010

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

INPUT ALPH. ⊂ "TABLE" ALPH.

EX. IN LECTURE NOTES : INPUT ALPH. = { ALL BYTES }
256 SYMBOLS (0..255)

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



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"~~ "" (EMPTY)
OR AN EXISTING TABLE ENTRY
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"

INPUT	S	TBL. INQ.		OUT	TABLE UPD.	
0	—	0	0	—	—	—
0	0	00	—	0	00	2
0	0	00	2	—	—	—
0	2	000	—	2	000	3
0	0	00	2	—	—	—
0	2	000	3	—	—	—
0	3	0000	—	3	0000	4
0	0	00	2	—	—	—
0	2	000	3	—	—	—
0	3	0000	4	—	—	—
0	4	00000	—	4	00000	5
—	0	—	—	0	—	—

"COMPRESSED" STREAM: "02340"

(WE ACTUALLY HAVE NOT WON ANYTHING YET)

LZW DECODER

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

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

UPDATE OF STATE / OUTPUT:

OUTPUT: TABLE (LAST SYMBOL)

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

LAST = NEW

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

Ex. 2 INPUT ALPH. = {0, 1, ... 255} (bytes)

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

COMPRESSED STREAM = (5, 7, 7, 5, 259, 5)

ORIGINAL STREAM = ?

IN	LAST	OUT	TABLE UPD.
5	—	—	—
7	5	5	256 : (5,7)
7	7	7	257 : (7,7)
5	7	7	258 : (7,5)
259	5	5	259 : (5,5)
5	259	(5,5)	260 : (5,5,5)
—	5	5	—

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? (FOR SOME INPUT SEQUENCE?)

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 $8 \cdot 3841$ BITS. HENCE COMPRESSION RATE $\leq \frac{8 \cdot 3841}{12} \ll 5000$