

Source coding techniques:

3- Lempel-Ziv LZ 78 COMPRESSION TECHNIQUES:

In 1978, Jacob Ziv and Abraham Lempel wrote an article entitled “Compression of Individual Sequences via Variable Rate Coding” in the IEEE Transactions on Information Theory describing a compression algorithm known as the LZ 78 algorithm.

LZ78 inserts one- or multi-character, non-overlapping, distinct patterns of the message to be encoded in a Dictionary. The multi-character patterns are of the form: $C_0C_1 \dots C_{n-1}C_n$. The prefix of a pattern consists of all the pattern characters except the last: $C_0C_1 \dots C_{n-1}$,

LZ78 Output:

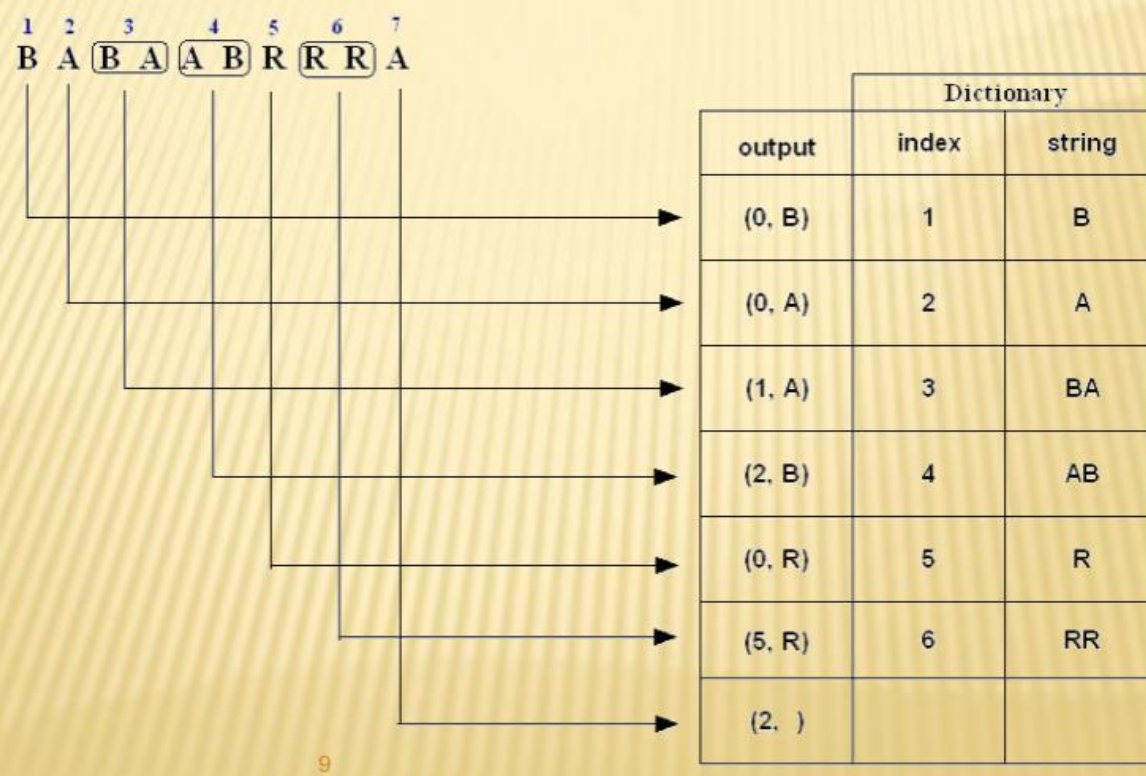
$(0, \text{char})$	if one-character pattern is not in Dictionary.
$(\text{DictionaryPrefixIndex}, \text{lastPatternCharacter})$	if multi-character pattern is not in Dictionary.
$(\text{DictionaryPrefixIndex},)$	if the last input character or the last pattern is in the Dictionary.

LZ78 COMPRESSION ALGORITHM

```
Dictionary ← empty; Prefix ← empty; DictionaryIndex ← 1;
While (characterStream is not empty)
{
    Char ← next character in characterStream;
    if(Prefix + Char exists in the Dictionary)
        Prefix ← Prefix + Char;
    else
    {
        if(Prefix is empty)
            CodeWordForPrefix ← 0;
        else
            CodeWordForPrefix ← DictionaryIndex for Prefix;
        Output: (CodeWordForPrefix, Char);
        insertInDictionary( ( DictionaryIndex, Prefix + Char) );
        DictionaryIndex++;
        Prefix ← empty;
    }
}
if(Prefix is not empty)
{
    CodeWordForPrefix ← DictionaryIndex for Prefix;
    Output: (CodeWordForPrefix, );
}
```

Example 1:

Encode (i.e., compress) the string **BABAABRRRA** using the LZ78 algorithm.



The compressed message is: (0,B)(0,A)(1,A)(2,B)(0,R)(5,R)(2,)

1. **B** is not in the Dictionary; insert it
2. **A** is not in the Dictionary; insert it
3. **B** is in the Dictionary.
BA is not in the Dictionary; insert it.
4. **A** is in the Dictionary.
AB is not in the Dictionary; insert it.
5. **R** is not in the Dictionary; insert it.
6. **R** is in the Dictionary.
RR is not in the Dictionary; insert it.
7. **A** is in the Dictionary and it is the last input character; output a pair containing its index: (2,)

Example 2:

Encode (i.e., compress) the string **AAAAAAAA** using the LZ78 algorithm.

1 2 3 4
A A A A A A A A A

Dictionary		
output	index	string
(0, A)	1	A
(1, A)	2	AA
(2, A)	3	AAA
(3,)		

1. A is not in the Dictionary; insert it
2. A is in the Dictionary
AA is not in the Dictionary; insert it
3. A is in the Dictionary.
AA is in the Dictionary.
AAA is not in the Dictionary; insert it.
4. A is in the Dictionary.
AA is in the Dictionary.
AAA is in the Dictionary and it is the last pattern; output a pair containing its index:
(3,)