RLE (Run Length Encoding)

RLE is the basic compression algorithm, it is the first compression ever invented (even though I have no prove nor document about it), as far as my research goes, RLE is patent free, which mean anyone can use it freely.

This compression method is so simple it just removing repetition of the same byte IN A ROW (continuously). This method is very important and all the compression utility should have this compression method built-in, I know that most of the compression utility have this method, this method is the first priority, it should be done before applying other compression method.

Eg: Source: ABCCCBBAAABBB
Encoding:

Input symbol	Output symbol	Total
A	A	1
B	B	1
CCC	C	3
BB	B	2
AAA	A	3
BBB	B	3

Result: A1B1C3B2A3B3

Decoding:

Input symbol	Total	Output symbol
A	1	A
B	1	B
C	3	CCC
B	2	BB
A	3	AAA
B	3	BBB

Result: ABCCCBBAAABBB

Note: this example shows every byte value followed by a number indicating how many repetition of the same byte value. This is a basic idea of RLE compression, it is not going to have a good result and instead of making the file smaller this method will make the file bigger, so this is not good but the example below shows the proper implementation, set minimum repeatition to 2 and it will compress only if two or more bytes are the same:

Encoding: search two byte of the same value and put a value of how many byte of the same value occur again, even no more is found it is still required to put zero value indicating no more is found.

Input symbol	Output symbol	Total
A	A	(none)
B	B	(none)
CCC	CC	1
BB	BB	0
AAA	AA	1
BBB	BB	1

Result: ABCC(1)BB(0)AA(1)BB(1).

Decoding: whenever we found two same charactera in a row it means the next byte value is the value of how many more repeatition occur.

Input symbol	Total	Output symbol
A	(none)	A
B	(none)	B
CC	1	CCC
BB	0	BB
AA	1	AAA
BB	1	BBB

Result: ABCCCBBAAABBB

My implemetation and source for RLE can be found in tools section, please download it and play with it, but remember in my RLE program I used the value of 4 instead of 2 (example above) as the minimum repeatition, I do this because the minimum value of 4 have a better result (normally).

Pros: simple, easy to coding and fast.
Cons: only able to compress if the characters are in a row.
Conclusion: this method is the simplest compression ever, its best use is for database file and still there are many ways to improve this method.

NEXT Arithmetic