

# Prétraitements

Prétraitement réversible des données avant d'appliquer le compresseur

amélioration : vitesse et/ou consommation mémoire et/ou taux de compression

- ▶ Run Length encoding (très simple) : contraction de séquences constantes
- ▶ Move to Front (simple) : densifier les codes de faible valeur
- ▶ Burrows-Wheeler (complexe) : faire apparaître des séquences constantes plus longues

# Run Length Encoding (RLE)

Une méthode de compression simple qui n'agit que sur les de suites de symboles identiques :

- ▶ plages de 0 dans section data (.o , exécutables),
- ▶ zones d'image de couleur uniforme.

▶ Seul : f rle f.z elr f

▶ Prétraitement : f rle lzw f.z wzl elr f

Ajouter une marque de répétition qui sera suivie (dans les données compressées) d'un nombre de répétitions :

$x \ x \ x \ \dots \ x \ x \ (n \text{ fois}) \rightarrow \text{rep}(x,n)$

- ▶ Code spécial disponible (ex : rep = 0 ou EOT en ASCII) :
  - ▶  $x \ x \ x \ x \ x \ x \rightarrow x \text{ Rep } 5$
  - ▶  $x \ x \rightarrow x \text{ Rep } 1$
- ▶ Répétition implicite sur double symbole :
  - ▶  $z \ x \ x \ \underline{x \ x \ x \ x} \ y \rightarrow z \ x \ x \ 4 \ y$
  - ▶  $z \ x \ x \ y \rightarrow z \ x \ x \ 0 \ y$  (*aucun y proche*)
  - ▶  $\underline{y \ z \ x \ x} \ y \rightarrow y \ z \ x \ x \ -3$  (delta codable vers y déjà traité).

# Move To Front

Après chaque lecture d'un symbole  $s$  le codage des symboles est mis à jour. Les symboles de codes inférieurs au code actuel de  $s$  sont décalés d'un cran et  $s$  est déplacé en tête (code 0).

Exemple : la chaîne "klklkabbbaaxxxzxz" comprend 3 portions comprenant chacune une répétition de paires de symboles.

Sans MTF : fréquence égale ( $1/6 \simeq 17\%$ ) des codes des 6 lettres.

Avec MTF : les lettres soulignées sont recodées 0 ou 1 (0 et 1 représentent à eux deux  $2/3 \simeq 67\%$  des symboles à compresser).

## Exemple MTF : "fcbbeghf"

Alphabet de 8 symboles codés sur 3 bits : a à h									
Codes binaires	Deux derniers symboles lus								
	" "	"f"	"fc"	"cb"	"bb"	"be"	"eg"	"gh"	"hf"
000	a	f	c	b	b	e	g	h	f
001	b	A	F	C	c	B	E	G	H
010	c	B	A	F	f	C	B	E	G
011	d	C	B	A	a	F	C	B	E
100	e	D	d	d	d	A	F	C	B
101	f	E	e	e	e	D	A	F	C
110	g	g	g	g	g	g	D	A	a
111	h	h	h	h	h	h	h	D	d

f      c      b      b      e      g      h      f  
 101 010 001 001 100 110 111 101    codage initial  
 101 011 011 000 101 110 111 101    codage mtf