

L'algo de compression

Cette présentation se donne comme objectif d'expliquer le principe de base d'un algorithme de compression.

Partant du constat que pour une taille donnée de fichiers l'entropie d'une séquence est inversement proportionnelle à sa longueur

Dit différemment :

si la taille d'un fichier est plus petit que toutes les combinaisons possibles d'une chaîne donnée il y a de quoi à faire

Partant de ce constat

2 données consécutives dans un fichier donc 16 bits

$$d = x_1 * 256 + x_2$$

maintenant je restreins volontairement x_1

x_1 compris entre 0 et 63

la valeur max donne donc $63 * 256 + 255 = 16383$ 11111111111111 14 bits

puis je rajoute un bit

pour savoir si j'ai lu le plus grand en premier ce qui va me donner le bits 15

donc avec ce codage je peux coder tous les couples (0..63) (0...255)

et ce quel que soit le sens dans le fichier à compresser

maintenant je rajoute un en-tête

qui va me donner les infos suivantes

nombre de pas de 63 $256/63=5$

quantité d'octets à traiter

les données qui suivent sont elles codées

et tout cela sur 16 bits

structure de l'en-tête

$2^{15}=0$ donnée non codée

$2^{14} \dots 2^0$ quantité d'octets à lire et à recopier sans décompression

$2^{15}=1$ donnée codée

$2^{13} \dots 2^3$ quantité d'octets à lire et à décompresser

$2^2 \dots 2^0$ nombre de décalage d'un pas de 63

pour le décodage

récupération de 15 bits

extraction du 15 bits

$d \% 256 = x_2$

$d / 256 + n * 63 = x_1$

et en fonction de la valeur du 15 bits j' écris x_0 et x_1 dans le bon sens

il est bien évident que pour que cela soit intéressant il faut faire un codage que si le gain est supérieur à 4 octets 2 en-tête

avec ce codage l'on a un bon petit gain pas génial mais il y a bien un gain

tout cela à cause des premières lignes de ce document

ce qui change des algos traditionnels, ceux qui font de la compression sans perte,

qui sont plus ou moins rangés dans la catégorie des algos à dictionnaire,

en gros l'on cherche une occurrence et l'on remplace cette

occurrence par une autre chaîne plus courte, l'un n'empêchant pas l'autre bien entendu

merci de votre attention remy aumeunier