

Codage

Compression

Liens :

Codage > Formats
Codage > Fichiers
Codage > Formats d'image
Codage > Son et vidéo

- Je dois envoyer un fichier par email mais sa taille est trop importante. Comment faire ?
- J'ai reçu un email avec un fichier ZIP en pièce jointe. Comment récupérer les informations qu'il contient ?
- Pourquoi la taille du fichier ne diminue-t-elle pas ou peu lorsque je « zippe » une vidéo ?
- Pourquoi y a-t-il une perte de qualité lorsque j'augmente le taux de compression d'une image au format JPEG ?

1. Principe général de la compression

Les données (texte, son, images, vidéo, programmes ou autres fichiers) ont souvent besoin d'être compressées pour des raisons d'espace de stockage et parfois de vitesse de transfert.

La compression consiste à réduire la taille d'un fichier numérique (cf. *Fichiers*) en codant dans un autre format les données qu'il contient.

Il s'agit donc de convertir des données d'un format d'origine vers un format compressé (cf. *Formats*), ce qui peut se faire **avec** perte ou **sans** perte d'information.

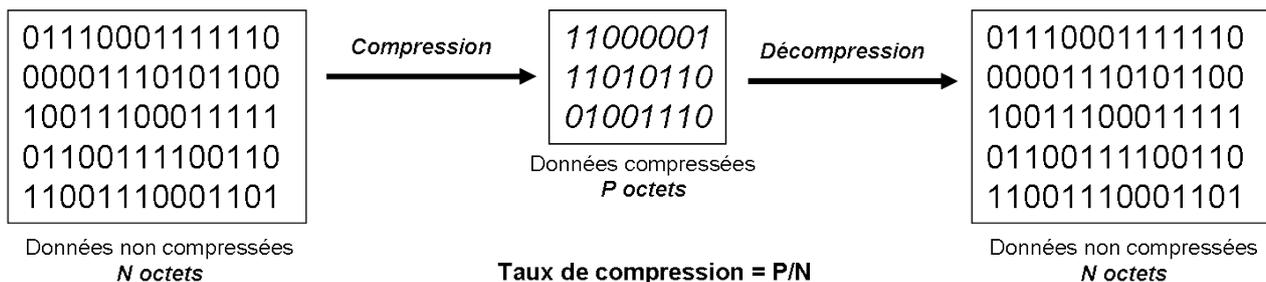


Figure 1 – Principe de la compression (cas d'une compression sans pertes)

L'efficacité de la compression d'un fichier est mesurée par le **taux de compression**, défini comme le rapport entre la taille du fichier compressé et celle du fichier initial. Avec cette définition, plus le taux de compression est faible, meilleure est la compression.

2. Compression d'un ou plusieurs fichiers : les archives

De nombreuses applications permettent de regrouper un ou plusieurs fichiers de formats divers ainsi que leurs chemins d'accès (cf. *Gestion des fichiers*) en un seul fichier compressé, à l'aide d'une méthode de compression **sans perte**. Un tel fichier compressé est fréquemment appelé **archive**.

Les formats d'archive les plus répandus sont **zip**, **rar**, **ace** et **tar.gz**.

Généralement ces applications prennent en charge les étapes de compression et de décompression. Par exemple, une application comme Power Archiver permet de compresser et décompresser des données en utilisant divers formats d'archive.



Remarque : La compression de fichiers déjà compressés est peu ou pas efficace. En effet le recodage des données qui pourrait aboutir à un gain de place a déjà été effectué lors de la première étape de compression.



Une archive n'est pas nécessairement un fichier compressé : il s'agit simplement d'un fichier dans lequel peuvent être regroupés plusieurs fichiers de tous types et leur chemin d'accès. Ainsi le format **tar** est un format d'archive non compressé. Cependant la plupart des formats d'archive (notamment **zip**, **rar**, **arj** ou encore **ace**) sont également des formats compressés.

3. Méthodes de compression sans perte

On appelle compression **sans perte** les méthodes qui permettent, après l'étape de décompression, de récupérer les données d'origine telles qu'elles étaient initialement.

La plupart sont basées sur des formules mathématiques qui permettent de traiter les redondances (c'est-à-dire les éléments qui apparaissent plusieurs fois à l'identique) en les remplaçant par des codes plus courts. Généralement ces méthodes sont indépendantes du type de données traitées.



La méthode de compression sans perte la plus simple est la compression RLE, utilisée par exemple dans les images au format **bmp**, qui consiste en un recodage des éléments identiques consécutifs.

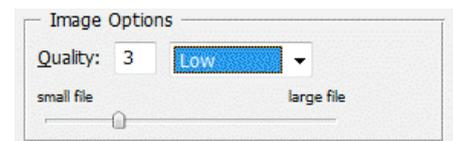
La méthode LZW (Lempel-Ziv-Welch) est utilisée dans les images aux formats **gif** et **tiff**, et est à la base de la compression appliquée aux archive **zip**, **rar** et **arj**. Le principe est de repérer des séquences qui apparaissent plusieurs fois, en construisant au fur et à mesure un dictionnaire de séquences, et de les remplacer par leur indice dans le dictionnaire. Le dictionnaire fait partie intégrante des données compressées. Pour que cette méthode soit efficace il ne faut donc pas que sa taille soit supérieure à celle économisée en recodant les données.

Il existe également des méthodes de compression sans perte dépendantes du type de données à compresser, qui tirent parti de connaissances a priori sur la structure de ces données. Par exemple, on peut améliorer les performances de compression d'un texte en utilisant la fréquence d'apparition des mots dans la langue concernée et en attribuant les codes les plus courts aux mots les plus fréquents, selon les principes du codage de Huffman.

4. Méthodes de compression avec perte

Les méthodes de compression **avec perte** s'appliquent à des types de données bien particuliers comme les images, les sons et les vidéos (cf. *Formats d'image, Son et vidéo*) et permettent généralement un gain de taille nettement supérieur à celui d'une compression sans pertes. Le principe, au-delà du recodage des redondances, est d'éliminer les informations les moins perceptibles par l'humain de manière à obtenir un résultat qui paraisse suffisamment proche de l'original après décompression.

Dans le cas d'une compression avec perte il faut trouver un compromis entre gain d'espace de stockage du fichier compressé et perte de qualité : en fonction de l'utilisation prévue on pourra se permettre de dégrader plus ou moins la qualité initiale des données en jouant sur le taux de compression. Ainsi dans le cas d'une image au format **jpeg** destinée à illustrer une page Web et devant se charger rapidement on choisira un taux de compression beaucoup plus élevé que pour une image destinée à des travaux d'impression de qualité professionnelle. La taille du fichier compressé sera donc nettement réduite mais un zoom sur l'image révélera une perte de qualité importante (cf. *Formats d'image*).



Dans le cas d'une compression avec perte les étapes de compression et de décompression sont plus souvent prises en charge par des applications distinctes. Ainsi, les programmes de lecture de médias comme Winamp permettent de lire des sons compressés aux formats **mp3**, **ogg** ou **wma** (cf. *Son et vidéo*), ce qui passe par une étape de décompression des données même si on n'enregistre pas de fichier décompressé, mais pas de créer ces fichiers compressés à partir de données brutes. A l'inverse une application comme CDex est exclusivement dédiée à l'**encodage** des données dans ces formats.

Références

- Compression : http://fr.wikipedia.org/wiki/Compression_de_donn%C3%A9es
- Archive : http://fr.wikipedia.org/wiki/Archive_%28informatique%29

Terminologie

- **Compression sans pertes** = compactage
- **Encodage** = compression avec pertes (vidéo)