Codage

Liens :
Codage > Codage numérique
Codage > Fichiers
Codage > Formats de texte
Codage > Formats d'image
Codage > Son et vidéo

Formats

- Comment une application fait-elle pour interpréter les données d'un document ?
- Comment transmettre une image créée avec Photoshop à quelqu'un qui n'a pas le logiciel Photoshop ?
- Pourquoi n'est-il pas possible de lire un document Word avec un éditeur de texte quelconque ?

1. Structuration des données informatiques

Les données informatiques sont représentées sous la forme d'une suite de 0 et de 1, quelque soit le type de fichier considéré (cf. *Codage numérique*, *Fichiers*). Une même suite de bits pourra donc avoir de multiples significations selon la façon dont elle est structurée.
Sans information sur la façon dont est organisée la suite de bits constituant les données, il n'est donc pas possible de décider a priori de la façon de les interpréter. De même, une application n'intégrant pas ces informations ne sera pas capable de créer un document dans lequel les données seront structurées selon les mêmes règles.

Pour la lecture comme pour l'écriture de données, il est nécessaire de connaître le langage, appelé **format**, utilisé pour décrire les informations manipulées sous la forme d'une suite de bits (cf. *Codage numérique*).

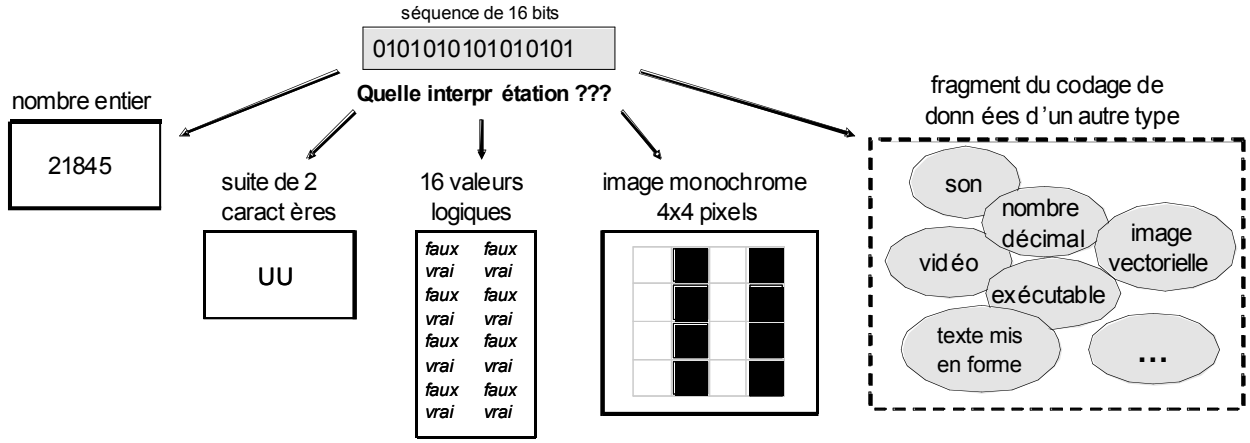


Figure 1 : De multiples interprétations possibles pour une même séquence de bits

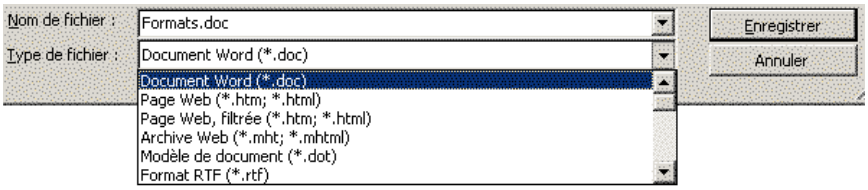
Un format est donc une convention, liée à l'usage qui est fait des données représentées dans le document, qui utilise souvent une combinaison d'éléments de base tels que ceux présentés ci-dessous. Par exemple le format DOC, qui est celui des documents Word, permet de coder à la fois un contenu textuel, des objets tels que des images, une mise en page et une mise en forme (cf. *Formats de texte*). Il existe souvent de nombreux formats destinés à coder le même type de données, sans qu'un format ne puisse être considéré dans l'absolu comme meilleur qu'un autre : une image peut ainsi être codée au format JPEG ou PNG, selon qu'on privilégie le gain d'espace de stockage ou la qualité de l'image (cf. *Formats d'image*). D'autre part différentes applications qui manipulent un même type de données développent souvent des formats spécifiques, non compatibles entre eux mais plus ou moins équivalents. C'est par exemple le cas des applications de dessin Gimp, Photoshop et PaintShop Pro dont le format par défaut permet de coder des images vectorielles et matricielles ainsi que des calques.

Le choix d'un format n'est pas guidé uniquement par la nature des données à coder, mais également par l'usage qui doit être fait de ces données.

Les systèmes d'exploitation établissent une correspondance entre type de document et application par défaut utilisée pour les traiter (cf. *Fichiers*). Cependant il n'y a pas de correspondance directe entre type de document et format, ne serait-ce que parce que les versions des formats évoluent au cours du temps. Afin que l'application utilisée pour ouvrir le document puisse interpréter les données de façon appropriée, de nombreux formats incluent un **en-tête** qui précise les caractéristiques de l'organisation des données : par exemple dans le cas d'une image bitmap, le nombre de bits utilisé pour coder la couleur de chaque pixel (cf. *Formats d'image*) est indiqué dans l'en-tête du fichier.

2. Conversion de formats

Pour convertir des données d'un format vers un autre, il est nécessaire de disposer d'une application capable à la fois de lire le format d'origine et de recoder les données en utilisant le format de destination. Dans la plupart des applications cela se fait par l'intermédiaire du menu *Enregistrer sous*. Par exemple, l'application Open Office Writer permet de convertir les informations codées dans un document au format DOC vers le format équivalent ODT.



Selon les caractéristiques de ces 2 formats, la conversion peut résulter en une perte d'une partie des données. Ainsi, l'application de traitement d'images *Paint* permet de convertir une image au format PNG vers le format GIF, moyennant une perte de qualité du codage des couleurs (cf. Images), mais ne prend pas en charge le format PSD spécifique à l'application Adobe Photoshop.

3. Pour aller plus loin : format ouvert et format fermé

Lorsque les caractéristiques d'un format sont diffusées librement et que ce format peut donc être mis en œuvre par n'importe quelle application et n'importe quel système d'exploitation (on parle alors d'**interopérabilité**), il s'agit d'un **format ouvert**. Par exemple les formats texte RTF et HTML, le format d'image PNG ou encore le format audio compressé OGG sont des formats ouverts.

A l'inverse les caractéristiques de certains formats, développés par des entreprises privées, sont tenues secrètes et/ou limitées dans leur utilisation par le cadre légal afin que ces formats ne puissent être traités que par certaines applications. On parle alors de **format fermé** ou de format propriétaire. C'est par exemple le cas des formats spécifiques aux applications Microsoft Office (DOC, XLS, PPT).

Lorsqu'un format n'est pas protégé par un brevet et que toute application peut donc l'utiliser sans avoir à payer de droits d'auteur, on parle de **format libre**, par opposition à un **format propriétaire**. Si les notions de format ouvert et format libre sont souvent confondues, de même que les notions de format fermé et format propriétaire, elles ne sont pas à proprement parler équivalente : par exemple le format PDF (cf. *Formats de texte*) est un format ouvert mais propriétaire, en effet ses caractéristiques sont publiques mais la société Adobe se réserve le droit de prélever des droits d'auteurs auprès des éditeurs de logiciels qui l'utilisent. Ce principe a également été adopté par la société XXX, détentrice des droits sur le format de fichiers sons compressés MP3 (cf. *Son et vidéo*).

Références

- Format de données : http://fr.wikipedia.org/wiki/Format_de_donn%C3%A9es
- Format ouvert : http://fr.wikipedia.org/wiki/Format_ouvert

Terminologie

- **Format ouvert** ≈ format libre
- **Format fermé** ≈ format propriétaire