

Théorie générale sur les données :

Les données sont réparties en trois catégories :

- En ouvrant le journal, nous avons affaire à des **nouvelles**, nous ne consultons alors pas encore le contenu qu'elles contiennent.
- Si la nouvelle apporte quelque chose de significatif et d'utile, elle devient une **information**.
- Si, au contraire, elle ne présente ni nouveauté, ni intérêt, elle devient alors une **redondance**.

Leur structure, elle aussi, peut être analysée de trois manières différentes :

- Les données non structurées, comme le texte, qui peut être **formaté**, ou non formaté.
- Les données faiblement structurées, comme dans un tableau, dont on peut voir la structure ci-dessous. Les données **doivent** être structurées pour être traitée par une machine.
- Les données fortement structurées, comme dans une base de données.

Numéro de client	Nom	→ Méta Information
33	Terrieur	
44	Sokoloff	→ Ligne, Enregistrement
55	Surchat	
↓ Colonne	↓ Cellule, Champ	

Lorsque l'on traite des données, on commence par les **distinguer**, pour ce faire, on peut suivre le modèle dit du **TPCSQ** (1. Type, 2. Provenance, 3. Caractéristique, 4. Structure, 5. Qualité).

Ensuite, on les **prépare**. Sont-elles traitables par informatique ? Si non, comment les préparer à un tel traitement puis, comment les sélectionner et en tirer les résultats qui nous intéressent ?

Enfin, on les **évalue** et on choisit comment les représenter. Par exemple, si on les présente sous forme de graphique, il convient d'en sélectionner un adapté.

Dans le cadre du projet Eureka, nous avons suivi ce modèle à la lettre sans y prêter attention. Nous avons d'abord recherché et **distingué** des données que nous avons ensuite **préparées** sous forme de tableaux, puis **évaluées** afin d'en créer des diagrammes pertinents à présenter à la classe.

Les différents types de données :

Type	Exemple de moyens de création	Exemple d'utilisation	Extensions
Texte	Bloc-notes, Word, Wordpad	Traitement de texte, création de documents	.txt / .doc(x)
Présentations	PowerPoint, Pezzi, PowToon	Création de présentations	.ppt(x/m)
Tableaux	Excel, Google Drive, Libre Office	Création de tableaux automatisés	.xls(x/m)
Base de données	MySQL, Access, Oracle Database	Traitement de données massives et automatisées	.db / .mdb
Diagrammes	Visio, Xmind, Dia Diagram Editor	Création de petits diagrammes aux éléments interdépendants	.xml / .svg
Graphiques*	Paint, Photoshop, Gimp	Création d'une image	.jpg / .png
Audio	Audacity, Ableton Live, FL Studio	Ecouter de la musique	.mp3 / .wav
Vidéo	Win Movie Maker, Sony Vegas	Regarder une vidéo ou un film	.wma / .avi
Programmes	Visual Studio, Teamspeak, etc...exe / .bat

* Pour ceux qui hibernaient pendant les cours, on parle ici de fichiers graphiques et non pas de graphiques (diagrammes).

Pour créer des analyses à partir de tableaux, il est important de bien formater les données:

Tableau :

Prénom	Jean-Luc	→ Texte
Naissance	31.02.1942	→ Date
Num. client	42	→ Nombre
Cotisation	400.-	→ Montant

Rappel sur les différents types de données :

Images :

- Les deux types d'images sont les images **bitmap** (dimensionnées en pixels), et les images **vectérielles** (constituées de plusieurs objets décrits par leur position, couleur, forme, etc...).
- Il existe plusieurs types d'**encodages** de la couleur, le **RGB** étant le plus courant et simple. Pour chacune de ces couleurs, on a 256 (pour 24 bits, 8 rouges, 8 verts et 8 bleus) valeurs, qui vont définir la couleur finale.
Exemple et formats : (255, 0, 0) = rouge ; (100%, 100%, 0%) = jaune ; #FFFFFF (FF = 255 en hexadécimal) = blanc.
- Attention, la **définition** (exemple : 1'920*1'080 pixels) et la **résolution** (exemple : 300 dpi) sont deux choses différentes et **souvent confondues** !
- Attention également à la distinction entre la **résolution** d'une image, définie en **dpi** (dots per inch), et celle de la **résolution d'une image sur un écran**, exprimée en **ppi** (pixel per inch) et qui dépend non plus de l'image mais du nombre de pixels gérés par la carte graphique.
- La **taille** en cm d'une image se calcule ainsi : 2.54 [constante] * Taille [inch = px/ppi].
Exemple : une image large de 500 pixels sur un écran d'une résolution de 72ppi fera $2.54 * (500/72) = 17.6\text{cm}$ de largeur.
- Sa **dimension** en pixels se calcule ainsi : (taille [cm]/2.54) * résolution [ppi].
Exemple : une image haute de 10cm affichée avec une résolution de 300ppi fait $10/2.54 * 300 = 1'181\text{px}$ de haut.
- Son **volume** en bytes (ou octets) d'une image non compressée se calcule ainsi :
Largeur [px] * hauteur [px] * profondeur [bit/pixel].
Exemple : une image de 1024*768 pixels encodée en RGB 24 bits a un volume de $1'024*768*(24/8) = 2'359'296$ b(ytes) ou 2.25 MB ($2'359'296 / 1024$ (kilobytes) / 1024 (megabytes)).
- Voici un tableau récapitulant les principaux formats d'images :

Format	Description, caractéristiques	Compression des données	Nombre de couleurs et encodage	Transparence	Compatible Web	Support animations	Type
JPEG	Joint Photographic Experts Group Format d'image le plus utilisé dans le monde. En particulier en photographie numérique.	Oui, réglable. Possible d'avoir du JPEG sans perte mais avec compression limitée.	16 millions (8 bits de couleurs, 24 bits de données). ^{*1}	Non	Oui	Non	Bitmap
GIF	Graphics Interchange Format Surtout utilisé pour les images animées.	Oui, sans perte (LZW ^{*2}).	16 millions (24 bits)	Oui	Oui	Oui	Bitmap
PNG	Portable Network Graphics Fichier lourd mais avec le moins de pertes.	Oui, sans perte (Deflate). Optionnelle.	16 millions (8 bits de couleurs - 24 bits de données). Jusqu'à 48 bits.	Oui, variable	Oui, sur les navigateurs récents	Non (oui si MNG ^{*3}).	Bitmap
TIFF	Tagged Image File Format Utilisé notamment pour des images de graphiques, mais aussi pour le scan, fax, etc...	Oui (LZW, PackBits, CCITT, etc...).	16 millions (8 bits de couleurs - 24 bits de données). Jusqu'à 64 bits.	Oui, variable	Non	Non	Bitmap
SVG	Scalable Vector Graphics Format vectoriel.	Sans perte en compressant le code source XML.	16 millions (24 bits)	Oui, variable	Oui, sur les navigateurs récents	Oui	Bitmap
BMP	Fichier Bitmap simple Premier format d'image créé par IBM et Microsoft.	Pas compressé jusqu'aux images de 18 bits.	De 1 à 24 bits	Non	Non	Non	Vectériel

^{*1}Probablement pas à connaître, mais pour aller plus loin : https://en.wikipedia.org/wiki/8-bit_color

^{*2}LZW, Deflate, PackBits, CCITT sont des algorithmes de compressions de données mais ce n'est pas à connaître non plus.

^{*3}MNG est une évolution du PNG supportant les animations, le PNG ne les supporte donc théoriquement pas.

Audio :

- La **fréquence du son**, exprimée en **Hz** (hertz), définit sa hauteur. La plage communément admise de sons audibles par l'homme va de **16 Hz** (en-dessous, on passe dans les infrasons) à **18-20'000 Hz** (au-dessus, on passe dans les ultrasons).
- Le **débit** d'un fichier audio, communément appelé **bitrate**, représente le flux de données de fichier. Il est donc le principal facteur de la qualité du son, mais aussi du poids du fichier de ce son. Il est exprimé en **kbps** ou **kbit/s** (kilobytes par seconde).
- A la page d'après se trouve un tableau récapitulant les principaux formats audio :

Format	Description, caractéristiques	Avantages	Inconvénients	Détails/Remarques	Lecteur
WAV	Waveform Audio File Format Standard de Microsoft et IBM.	Aucune perte, reconnu sous Linux et Mac.	Grande taille (> 2GB), pas de tag.	Pas compressé, remplacé par le bwf.	Tous
CDA	Compact Disk Audio Raccourci créé par Windows lors de la lecture d'un CD.	-	Inutile sans insérer CD dans le lecteur.	Ne contient que des infos sur la piste à lire.	Tous
MID(I)	Musical Instrument Digital (Interface) Format de fichiers audio ne convenant qu'à des sons créés et reproductibles par un synthétiseur.	Taille très petite.	Mauvaise qualité, juste pour musique électronique.	Ne contient que des notes.	Lecteurs midi, synthétiseurs, WMP.
MP3	MPEG-1/2 Audio Layer 3 Fichier audio compressé, très répandu pour l'écoute de musique.	Compression bonne et modulable, bonne qualité, gère les tags, pas protégé.	Compression avec pertes, pas libre de droits, que 2 canaux.	Suppression des fréquences <20Hz, >20Khz.	Lecteurs MP3, WMP, lecteurs avec codec MP3, VLC, ...
WMA	Windows Media Audio Media de Microsoft.	Plus petit que WAV, gère les tags et les DRM, meilleure qualité que MP3.	Perte de qualité.	Différents types : standard, pro, loseless. Qualité variable.	WMP, VLC.
Ogg	« Ogging » Fichier audio plus performant que le MP3.	Logiciel libre, polycanaux, meilleure compression que MP3	Trop peu répandu, compression destructive.	-	VLC, WMP, Sonic, iTunes.

Vidéo :

- Voici un tableau récapitulant les principaux formats vidéo :

Format	Description, caractéristiques	Détails, remarques
WMV	Windows Media Video Contenu développé par Microsoft destiné à véhiculer des données audio et vidéo compressées (avec pertes).	Nombreuses versions : HD pour « haute définition », AP pour « advanced profile » (optimise la diffusion du flux vidéo).
MOV	QuickTime Movie QuickTime est le nom de la technologie multimédia d'Apple, créée en 1989 puis mise sur le marché en 1991 pour le Macintosh. Elle fut portée sur Windows en 1992 afin d'accélérer son adoption par l'industrie.	Un fichier QuickTime contient une ou plusieurs pistes, chacune comportant un type de données particulier : audio, vidéo, effet ou texte (pour des sous-titres par exemple).
AVI	Audio Video Interleave Format de fichier conçu pour stocker des données audio et vidéo. Il a été présenté par Microsoft en nombre 1992, en tant qu'élément de la vidéo pour la technologie de Windows.	Dans un fichier AVI, chaque composante audio ou vidéo peut être compressée par n'importe quel codec. Le format DivX est souvent utilisé comme codec vidéo, et le format mp3 comme codec audio, mais d'autres codecs peuvent également être utilisés, par exemple XviD, MPEG pour la vidéo, etc.
MKV	Matroska Video Format de fichier libre et donc sans redevance. Matroska n'est pas un codec (comme MPEG-1, XviD ou DivX) mais un conteneur, au même titre que l'AVI.	Il supporte quasiment tous les flux vidéo, quasiment tous les flux audio (en nombre illimité) et il peut contenir un grand nombre de pistes de sous-titre de différents formats.
DivX	Aussi appelé MPEG4 ou MP4, le DivX est un format de compression qui permet, avec les codecs actuels, de stocker sur un CD de 700 Mo un film d'une durée de 1h30 avec une qualité d'image et de son très proche du DVD, et bien meilleure que les cassettes VHS.	Ce n'est pas un logiciel libre et son code source n'est pas disponible mais une version publique, appelée OpenDivX, a été mise à disposition par DivXNetworks au début de l'année 2001.