

Les types de fichiers

Tristan Crolard, Sovanna Tan

Septembre 2007, m.a.j. septembre 2014

disponible sur

http://lacl.fr/tan/L1/types_fichiers.pdf

- 1 Introduction
- 2 Types de fichiers
- 3 Fichiers textes
- 4 Fichiers du web
- 5 Bureautique
- 6 Archivage
- 7 Multimédia

Qu'est-ce qu'un fichier ?

Un fichier est une suite de 0 et de 1 représentant de l'information codée avec un format prédéfini.

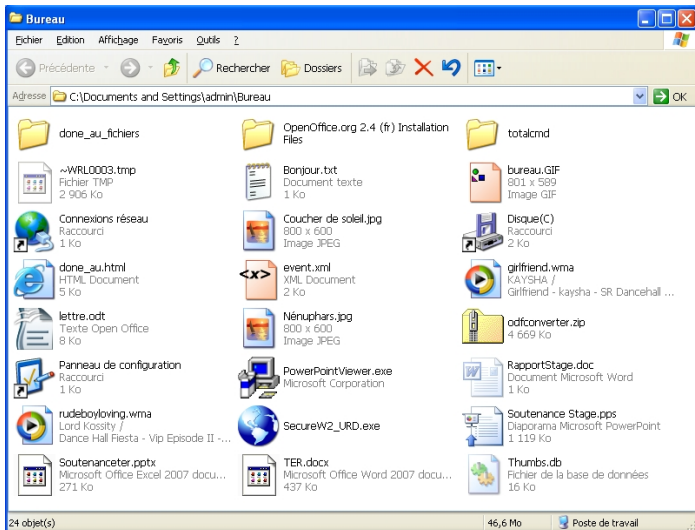
Un système d'exploitation d'ordinateur stocke et exploite les fichiers sur des supports variés : disque dur, clé usb, bande magnétique, cd, dvd, téléphone portable, lecteur mp3, ...

Différents types de documents

- Selon l'application
 - Bureautique
 - Internet
 - multimédia
- Pour la diffusion
 - par courrier électronique
 - sur un site web
- Pour l'archivage
 - formats d'archives
 - supports

... indépendamment du système d'exploitation.

Différents fichiers



Les programmes

Fichier exécutable

Fichier qui constitue un programme informatique, qui contient des instructions ou commandes exécutées par l'ordinateur.

Binaire

Fichier directement codé avec des instructions pour le processeur.

Script

Fichier texte contenant des commandes qui sont interprétées, traduites à la volée en instructions pour le processeur.

Le nommage des fichiers

nom.extension

Contraintes imposées par les systèmes d'exploitation

- 8 caractères pour le nom et 3 pour l'extension sous MS-DOS.
- 256 caractères pour le nom et l'extension sous Windows.
- 256 sous les systèmes Unix dont Linux, extension optionnelle de longueur variable.

- 1 Introduction
- 2 Types de fichiers
- 3 Fichiers textes
- 4 Fichiers du web
- 5 Bureautique
- 6 Archivage
- 7 Multimédia

Différents types de fichiers

- Les .txt sont des fichiers textes.
- Les .doc sont créés par le traitement de texte Microsoft Word.
- Les .xls sont créés par le tableur Microsoft Excel.
- Les .ppt sont créés par Microsoft Powerpoint.
- Les .pdf sont lisibles par Adobe Reader.
- Les .bmp, .gif, .jpg, ... sont des fichiers d'images.
- Les .mov, .avi, .mpg, ... sont des vidéos.
- Les .wav, .aif, .mp3 ... sont des fichiers de sons.
- Les .htm, .xml ... sont les fichiers diffusés par les serveurs Web.
- ...

Types MIME

L'acronyme MIME signifie Multipurpose Internet Mail Extensions.

L'organisme international IANA (Internet Assigned Numbers Authority) gère les types de documents déclarés.

La correspondance se configure directement sur l'ordinateur.

Types MIME

Extension	Type MIME	Extension	Type MIME
.avi	video/x-msvideo	.mp3	audio/mpeg
.bin	application/octet-stream	.mpeg	video/mpeg
.c	text/plain	.mpg	video/mpeg
.css	text/css	.pdf	application/pdf
.doc	application/msword	.png	image/png
.eps	application/postscript	.ppt	application/mspowerpoint
.exe	application/octet-stream	.ps	application/postscript
.gif	image/gif	.rtf	text/rtf
.gz	application/x-gzip	.tar	application/x-tar
.htm	text/html	.txt	text/plain
.html	text/html	.xls	application/vnd.ms-excel
.jpe	image/jpeg	.xml	text/xml
.jpeg	image/jpeg	.zip	application/zip
.jpg	image/jpeg	...	
.mov	video/quicktime		

- 1 Introduction
- 2 Types de fichiers
- 3 Fichiers textes**
- 4 Fichiers du web
- 5 Bureautique
- 6 Archivage
- 7 Multimédia

Fichiers textes

Un fichier texte est un fichier composé de caractères stockés sous la forme d'octets.

- Un bit est un élément de $\{0, 1\}$.
- Un octet est une suite de 8 bits.

Différents codages

American Standard Code for Information Interchange (ASCII) :

Jeu de caractères anglo-saxon, sans accents.

ISO 8859-1 (aussi appelé ISO Latin 1) :

Jeu de caractères accentués de l'europe de l'ouest.

Unicode (standard ISO/IEC 10646)

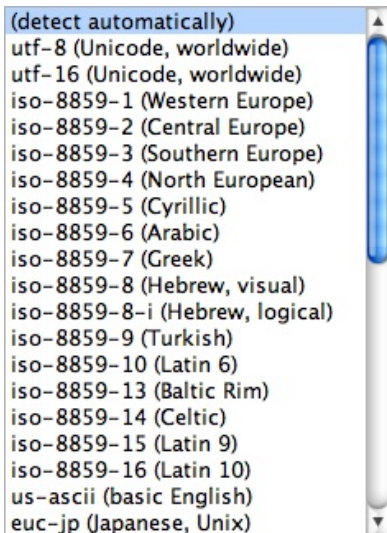
Tous les caractères sur 32 bits (encodage le plus fréquent UTF-8).

La famille ISO 8859

- Latin1 (West European)
- Latin2 (East European)
- Latin3 (South European)
- Latin4 (North European)
- Cyrillic
- Arabic
- Greek
- Hebrew
- Latin5 (Turkish)
- Latin6 (Nordic)
- ...
- Latin9 (\approx Latin1 et le symbole "euro")

Les codes 0 – 127 contiennent toujours les caractères ASCII.

Encodage des caractères



- 1 Introduction
- 2 Types de fichiers
- 3 Fichiers textes
- 4 Fichiers du web**
- 5 Bureautique
- 6 Archivage
- 7 Multimédia

Les documents sur le web

HTML (4.01) :

HyperText Markup Language 1999

XHTML (1.1) :

Extensible HyperText Markup Language 2001

CSS (2) :

Cascading Style Sheets 1998

XML (1.1) :

Extensible Markup Language 2004

XSL 1.0 :

Extensible Stylesheet Language 2001

Les documents sur le web

(detect automatically)
XHTML 1.1
XHTML Basic 1.0
XHTML 1.0 Strict
XHTML 1.0 Transitional
XHTML 1.0 Frameset
ISO/IEC 15445:2000 ("ISO HTML")
HTML 4.01 Strict
HTML 4.01 Transitional
HTML 4.01 Frameset
HTML 3.2
HTML 2.0
MathML 2.0
XHTML 1.1 plus MathML 2.0

Exemple de fichier XHTML

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <meta http-equiv="Content-Type"
      content="text/html; charset=UTF-8" />
    <meta http-equiv="Content-Style-Type" content="text/css" />
    <title>Exemple</title>
  </head>
  <body>
    <h1>
      <b>Mon titre</b>
    </h1>
    <p>Mon texte...</p>
  </body>
</html>
```

Exemple de fichier XML

```
<?xml version="1.0" encoding="ISO8859-1" ?>
<?xml-stylesheet type="text/css" href="cd\_catalog.css"?>
<CATALOG>
<CD>
<TITLE>Empire Burlesque</TITLE>
<ARTIST>Bob Dylan</ARTIST>
<COUNTRY>USA</COUNTRY>
<COMPANY>Columbia</COMPANY>
<YEAR>1985</YEAR>
</CD>
<CD>
<TITLE>Hide your heart</TITLE>
<ARTIST>Bonnie Tylor</ARTIST>
<COUNTRY>UK</COUNTRY>
<COMPANY>CBS Records</COMPANY>
<YEAR>1988</YEAR>
</CD>
</CATALOG>
```

Exemple de fichier CSS

```
CATALOG {
background-color: #ffffff;
width: 100\%;
}
CD {
display: block;
margin-bottom: 30pt;
}
TITLE {
font-size: 20pt;
}
ARTIST {
color: #0000FF;
}
COUNTRY,PRICE,YEAR,COMPANY {
display: block;
margin-left: 20pt;
}
```

Les flux RSS

Really Simple Syndication

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="updates.xslt" ?>
<rss version="2.0" xmlns:atom="http://www.w3.org/2005/Atom">
  <channel>
    <item>
      <title>Java</title>
      <link>http://lacl.univ-paris12.fr/tan/Java/java.html</link>
      <description>Du deuxième au cinquième TP ... </description>
    </item>
    <item>
      <title>Php</title>
      <link>
        http://lacl.univ-paris12.fr/verlan/enseignement/PWEB/index.html
      </link>
      <description>Le deuxième TP ...</description>
    </item>
  </channel>
</rss>
```


- 1 Introduction
- 2 Types de fichiers
- 3 Fichiers textes
- 4 Fichiers du web
- 5 Bureautique**
- 6 Archivage
- 7 Multimédia

Bureautique

Format standard pour les documents :

- OpenDocument 1.0 (ISO/IEC 26300 depuis mai 2006) basé sur XML.
- Office Open XML ou OOXML (ISO/IEC DIS 29500 depuis mars 2008). Le XML de Microsoft est devenu une norme.

Les applications les plus courantes

- OpenOffice/StarOffice (Sun) (*compatible OpenDocument*)
- KOffice (KDE) (*compatible OpenDocument*)
- Lotus Notes (IBM) (*compatible OpenDocument*)
- Office (Microsoft) (*compatible Office Open XML*)
 - Format ouvert : Rich Text Format (RTF)

Les standards bureautique

Fichier	Extension	Type Mime
Text	.odt	application/vnd.oasis.opendocument.text
Spreadsheet	.ods	application/vnd.oasis.opendocument.spreadsheet
Presentation	.odp	application/vnd.oasis.opendocument.presentation
Drawing	.odg	application/vnd.oasis.opendocument.graphics
Chart	.odc	application/vnd.oasis.opendocument.chart
Formula	.odf	application/vnd.oasis.opendocument.formula
Database	.odb	application/vnd.oasis.opendocument.database
Image	.odi	application/vnd.oasis.opendocument.image
Master Document	.odm	application/vnd.oasis.opendocument.text-master
Text Microsoft Word	.docx	application/vnd.openxmlformats
Microsoft Excel	.xlsx	application/vnd.openxmlformats
Microsoft PowerPoint	.pptx	application/vnd.openxmlformats

Exemple de document texte

Créer un fichier `essai.odt` contenant :

Mon titre

Mon texte...

Exemple de document texte

Le fichier `essai.odt` est en fait une archive zip contenant :

Length	Date	Time	Name
-----	----	----	----
39	09-25-05	10:12	mimetype
0	09-25-05	10:12	Configurations2/
0	09-25-05	10:12	Pictures/
2810	09-25-05	10:12	content.xml
8859	09-25-05	10:12	styles.xml
1054	09-25-05	10:12	meta.xml
6607	09-25-05	10:12	settings.xml
901	09-25-05	10:12	META-INF/manifest.xml
-----			-----
20270			8 files

Exemple : contenu de content.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<office:document-content ... >
  <office:font-face-decls>
    <style:font-face style:name="Lucidasans1" svg:font-family="Lucidasans"/>
    ...
  </office:font-face-decls>
  <office:body>
    <office:text>
      <text:h text:style-name="Heading_20_1" text:outline-level="1">
        Mon titre
      </text:h>
      <text:p text:style-name="Text_20_body">
        Mon texte...
      </text:p>
    </office:text>
  </office:body>
</office:document-content>
```

Format pour l'impression

PDF (Adobe)

L'acronyme PDF signifie Portable Document Format.

- Pour la visualisation ou l'impression de documents non éditables.
- Remplace le postscript (Adobe).
- Logiciel gratuit (non open source) pour visualiser et imprimer : Adobe Reader.
- Spécification "ouverte" de Adobe.
- Standards ISO PDF/X-1a :2003 et PDF/X-3 :2003 (uniquement pour sous-ensemble du PDF)
- Support natif sous Mac OS X.

- 1 Introduction
- 2 Types de fichiers
- 3 Fichiers textes
- 4 Fichiers du web
- 5 Bureautique
- 6 Archivage**
- 7 Multimédia

Archives compressées ou non

`.tar` :

Tape ARchive format defined by POSIX 1003.1-2001. specification.

`.gz` :

IETF RFC 1952 - GZIP file format specification version 4.3.

`.zip` :

Spécification "ouverte" de PKWARE (non standardisé).

`.rar` :

Roshal ARchive, format propriétaire inventé par Eugène Roshal, efficace pour la découpe des fichiers.

`.bz2` :

algorithme de compression très efficace utilisé pour la diffusion des logiciels UNIX.

- 1 Introduction
- 2 Types de fichiers
- 3 Fichiers textes
- 4 Fichiers du web
- 5 Bureautique
- 6 Archivage
- 7 Multimédia**

Images au format vectoriel

Image vectorielle

Format d'image où l'image est représentée par des objets mathématiques.

Exemple : cercle (centre, rayon)

- Scalable Vector Graphics (SVG du W3C)
- Adobe PDF (Acrobat)
- Adobe Illustrator
- Encapsuled postscript (EPS)
- QuarkXpress (QXD)
- Macromedia Flash (format d'animation vectorielle)
- Drawing eXchange Format (DXF d'Autocad)

Images matricielles

Image bitmap

tableau de pixels qui définit la couleur de chaque point de l'image

Portable Network Graphics (PNG ISO 15948) :

compression *sans perte*

Joint Photographic Expert Group (JPEG ISO 10918) :

compression *avec perte* (\Rightarrow meilleure compression possible)

JPEG-2000 (ISO/IEC 15444) :

compression avec ou sans perte

Video (standards ISO/IEC)

Différents CODEC (COmpression/DECompression)

MPEG-1 (Moving Picture Expert Group 1991)

qualité VHS

MPEG-2 (1994)

qualité DVD

MPEG-4 (1998)

- Basé sur le format de fichier d'Apple QuickTime
- Meilleure compression (équivalent au "DivX")

Audio

MP3 :

MPEG-1 layer III

AAC (Advanced Audio Coding) :

ajouté au MPEG-2 (repris dans MPEG-4)

Format de données 3D

- Virtual Reality Markup Language (VRML de Web3D)
- Extensible 3D (X3D de Web3D)
- Initial Graphics Exchange Specification (IGES, norme pour la CAO)
- BLEND (Blender)
- Drawing eXchange Format (DXF d'Autocad)

Merci.