



COMPLÉMENTS FORMATS DE FICHIERS, INTERNET, RECHERCHE D'INFORMATION

Stage – Semaine 5



RAPPEL : AU SEMESTRE 2 !

Au semestre 2, vous serez en niveau « Standard ». Ce niveau reprend le contenu du stage pour aller ensuite au-delà. En fonction de votre résultat, vous aurez 6 ou 10 semaines d'enseignement (en d'autres termes, vous serez dispensé ou non de la première partie du niveau).

- **Par défaut, l'enseignement dure tout le semestre.** Si vous êtes en « Standard 10 semaines », il n'y a pas de vagues d'enseignement.
- Si vous êtes en « Standard 6 semaines », vous restez dans la même vague qu'au premier semestre.
- A suivre sur le site d'enseignement
 - Rubrique « Résultats » pour connaître votre niveau au semestre 2
 - Rubrique « Calendrier » pour connaître les principales dates du niveau dans lequel vous serez.

<http://www.univ-montp3.fr/map/ens/info/>

2



ATTENTION !

Pour l'examen (15 points de la note finale), la semaine prochaine, vous devez :

- Avoir **votre carte d'étudiant**
- Connaître **votre mot de passe** pour accéder à **Moodle**

D'ici l'examen :

- Pensez à **terminer les tests de contrôle continu** (5 points de la note finale)

3



FORMATS DE FICHIERS

Stage – Semaine 5



FORMATS DE FICHIERS

- Déjà vu : il existe plusieurs formats d'images ?
 - à cause de choix de codage différents
 - Codage matriciel versus codage vectoriel
 - Codage de couleurs (niveaux de gris, modèle RVB, ...)
 - Utilisation ou non d'algorithme de compression (potentiellement destructrice)
- La multitude de formats existe pour d'autres types de documents numériques
 - La raison est toujours des codages différents...
 - ... en suite de 0 et de 1, le bit étant l'information de base pour un ordinateur



CODAGE ?

Codage d'une information :
Règle d'écriture à l'aide d'un nombre fini de symboles permettant de désigner de manière unique l'information (le codage est contextualisé (codage de nombre, de couleurs, de textes, ...))

Exemple : codage d'entiers

123 ↔ Cent vingt trois ↔ CXXIII

Trois écritures de la même valeur

à l'aide d'une liste de symboles indécomposables :

- des chiffres (0,1,2, ...)
- des mots (un, deux, dix, cinquante, ...)
- des lettres (I,V,X,L,C, ...)

6

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

PRINCIPAUX FORMATS POUR LES TEXTES

Trois niveaux de formats dépendant du type de contenu

- contenu seulement** (texte brut)
 - Format dit TEXTE : txt
 - Texte formaté par l'utilisateur : html, xml, ...
 - Programmes informatiques : php, c, ...
- les enrichissements** (police, corps, couleurs, ...)
- fonctionnalités avancées** (pagination, tables, index, ...)

Format TEXTE ENRICHIS: rtf

Format de travail de bureautique : doc, docx, odt

Format d'échange et de distribution : pdf

7

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

LE FORMAT TEXTE SEULEMENT

- La seule information qu'il contient = une suite de caractères
- Conseillé d'utiliser comme logiciel, un « éditeur de texte » (TextWrangler, BBedit, emacs, NotePad++, ...) et non un « logiciel de traitement de texte » (TextEdit, Open Office Writer, Libre Office Writer, Microsoft Word, ...)
- Chaque caractère est codé par (i.e. associé à) une suite de 0 et de 1

Exemple : codage de A et B en code ASCII ou UTF-8

A 0 1 0 0 0 0 0 1

B 0 1 0 0 0 0 1 0

- ... mais plusieurs tables de codage possibles
 - ASCII
 - Unicode (utf-8, ...)

(voir diapositive suivante)

8

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

DE L'ASCII À L'UTF-8

- ASCII d'origine** (aperçu à droite)
 - 7 bits par caractère : caractères latins sans accent.
 - 128 caractères
- ASCII étendu**
 - nombreuses extensions non compatibles entre elles (une par langue)
 - 8bits par caractère
 - 256 caractères
- « utf-8 » :
 - une table universelle (recommandée W3C)
 - Codage variant de 8 à 32 bits pour un caractère ou idéogramme

盲人摸象 - 各執一端

Traduction : máng rén mō xiàng... gè zhí yì duān

Caractère	Décimal Number	Binary Number	Caractère	Décimal Number	Binary Number
Null space	32	001010000	z	96	01101100
!	33	001010001	[95	01011111
"	34	001010010]	96	01101000
#	35	001010011	^	97	01101001
\$	36	001010100	_	98	011010010
%	37	001010101	`	99	011010100
&	38	001010110	a	100	011010101
'	39	001010111	b	101	011010110
(40	001011000	c	102	011010111
)	41	001011001	d	103	011011000
*	42	001011010	e	104	011011001
+	43	001011011	f	105	011011010
,	44	001011100	g	106	011011011
-	45	001011101	h	107	011011100
.	46	001011110	i	108	011011101
/	47	001011111	j	109	011011110
0	48	001100000	k	110	011011111
1	49	001100001	l	111	011100000
2	50	001100010	m	112	011100001
3	51	001100011	n	113	011100010
4	52	001100100	o	114	011100011
5	53	001100101	p	115	011100100
6	54	001100110	q	116	011100101
7	55	001100111	r	117	011100110
8	56	001101000	s	118	011100111
9	57	001101001	t	119	011101000
A	65	001101010	u	120	011101001
B	66	001101011	v	121	011101010
C	67	001101100	w	122	011101011
D	68	001101101	x	123	011101100
E	69	001101110	y	124	011101101
F	70	001101111	z	125	011101110
G	71	001110000	[126	011101111
H	72	001110001]	127	011110000
I	73	001110010	^	128	011110001
J	74	001110011	_	129	011110010
K	75	001110100	`	130	011110011
L	76	001110101	a	131	011110100
M	77	001110110	b	132	011110101
N	78	001110111	c	133	011110110
O	79	001111000	d	134	011110111
P	80	001111001	e	135	011111000
Q	81	001111010	f	136	011111001
R	82	001111011	g	137	011111010
S	83	001111010	h	138	011111011
T	84	001111100	i	139	011111100
U	85	001111101	j	140	011111101
V	86	001111110	k	141	011111110
W	87	001111111	l	142	011111111
X	88	001110000	m	143	011110000
Y	89	001110001	n	144	011110001
Z	90	001110010	o	145	011110010
[91	001110011	p	146	011110011
\	92	001110100	q	147	011110100
]	93	001110101	r	148	011110101

9

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

RECONNAISSANCE DE FORMATS

- Extension de fichiers** : suffixe de 2 à 4 lettres indiquant le format utilisé pour coder l'information contenue dans le fichier
 - Exemples : txt, odt, docx, pdf, ..., jpg (pour format jpeg), ...
- Le logiciel système associe un logiciel à chaque format (le logiciel associé par défaut peut-être modifié).
- Si le suffixe associé à un fichier est erroné (ne correspond pas au format réellement associé), le logiciel par défaut risque de ne pas savoir ouvrir le document → fichier inutilisable
- Conséquence : conseil = **laisser les logiciels ajouter eux-mêmes les extensions** (Ne saisissez que les noms hors suffixes) car ils connaissent l'extension adéquate pour le format qu'ils utilisent

10

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

TRANSFORMATION DE FORMATS

- Changer (ou saisir) un suffixe ne change pas le format réel du contenu d'un fichier.
- Par contre, les logiciels connaissent généralement :
 - Plusieurs formats
 - Les traitements pour passer de l'un à l'autre (via l'enregistrement ou l'export du fichier... en précisant le « type de fichier » ou le « format de fichier »)
- Ces fonctionnalités d'enregistrement ou d'export sont à utiliser pour changer le format d'un fichier.

11

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

INTERNET

Stage – Semaine 5

12

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

Est-ce qu'internet = web ?

- NON
 - Internet = réseau physique mondial d'ordinateurs (ou plus précisément réseau de réseaux)
 - Internet à 50 ans !
 - Le 29 octobre 1969, transfert de « lo » puis une heure après « login »
 - Voir : https://www.lk.cs.ucla.edu/internet_first_words.html
 - Web (World Wide Web) = réseau d'informations constitués des documents mis à disposition sur les serveurs d'internet et reliés par les hyperliens qu'ils contiennent.
 - Créé au début des années 1990

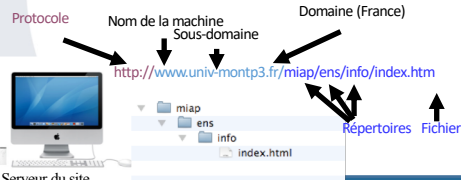
UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

RAPPELS : CLIENT-SERVEUR

- Principe de base :
 - Un ordinateur (le client) demande un service (exemple : une page web) à un serveur
 - L'échange est effectué en suivant un ou des **protocoles, langages** gérant le dialogue entre machines
- Quelques protocoles :
 - **IP** (Internet Protocol) : gestion d'internet
 - **http** (HyperText Transfer Protocol) : gestion web
 - **https** : idem sécurisé (crypté)
 - **imap** (Interactive Message Access Protocol) : accès à des messageries électroniques à partir de logiciels clients
 - **smtp** (Simple Mail Transfer Protocol) : envoi de mails
 - **ftp** (File Transfer Protocol) : transfert de fichiers (utile quand vous hébergez un site web chez un fournisseur de services)
- Remarque : pare-feu : permet de gérer les protocoles autorisées

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

ADRESSAGE

- Nécessité de localisation : chaque ordinateur est associée à une adresse l'identifiant gérée par le protocole IP
 - Adresse IP = suite de 32 bits (représentés sous forme de 4 blocs de nombres entre 0 et 255 (8 bits)) en IPv4 ou 128bits en IPv6 en cours de déploiement
 - Exemples : 193.52.137.213, 10.3.7.12 ...
- Pour le protocole http, localisation d'une page web :
 

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

SITE ?

- Un site :
 - Ensemble de pages web dont l'URL débute par un même début, appelé racine du site ou URL du site.
- Exemples :
 - <http://www.univ-montp3.fr/> : site de l'université
 - <http://www.univ-montp3.fr/miap/ens/info/index.htm> : site des enseignements d'informatique

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

TRACES

- Lors d'une navigation sur le web, les données échangées avec les serveurs interrogés passent par des serveurs intermédiaires
 - Les données peuvent être lues par d'autres.
 - Vos requêtes peuvent être stockées (obligation légale pendant 1 an des fournisseurs de services – LCEN : Loi pour la confiance dans l'économie numérique)
- Autres traces
 - Cookies (avec RGPD – Règlement général sur la protection des données – vous connaissez !)
 - Historique de navigation
 - Métadonnées dans les fichiers
 - ...

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

Que se passe-t-il quand on saisit l'adresse IP d'une machine au lieu de l'URL d'un site ?

- Si l'IP correspond à l'adresse d'un site web, l'adresse IP est remplacée par l'URL du site et la page d'accueil du site est affichée (d'autres mécanismes peuvent être mis en place).
- Exemple : 193.52.137.213 ... serveur web de l'université (mais des redirections qui mettent en alerte Firefox)

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

QUELS OUTILS UTILISEZ-VOUS LORS D'UNE RECHERCHE D'INFORMATION SUR LE WEB ?

1. Un ordinateur (de bureau, portable, smartphone, ...) connecté à internet
2. Un navigateur Web :
 - Chrome, Firefox, Edge, Safari, Opéra, ...
3. Un moteur de recherche : (site contenant un index d'une partie du web)
 - Google, Qwant, Bing, ...
 - ou un méta-moteur de recherche (site construisant sa réponse en recoupant les résultats de requêtes à différents moteurs de recherche)
 - DuckDuckGo, Lilo, ...

• Remarque : moteur de recherche ≠ navigateur web

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

EST-CE QU'UN MOTEUR DE RECHERCHE INDEXE TOUT LE WEB ?

- Non
 - Le web est dynamique :
 - Des pages se créent (et disparaissent) régulièrement
 - Les crawlers ou spiders, robots d'indexation, ne peuvent pas tout recenser en temps réel
 - Potentiellement un problème de place dans la base de données d'indexation (et donc des choix sur ce que conserver)
 - Des sites non reliés aux autres (web invisible)
 - Des pages payantes ou privatisées par mot de passe
 - ...

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

DANS LE TD

- Un zoom sur des requêtes dans un moteur de recherche
 - But = réduire le nombre de réponses et obtenir des réponses plus précises
- Réflexion sur la qualité des sites consultés :

La qualité des ressources est très variable.
C'est l'internaute qui doit la juger.