

HNU6051-A24

Humanités numériques : formats d'écriture

Mathilde Verstraete

Automne 2024

Contenu

Bienvenue	5
Support de cours	5
Crédits	5
Plan de cours	6
Informations pratiques	6
Description et objectifs de l'atelier	6
Support de cours	6
Organisation des séances	7
Séance 1 (4 octobre) :	7
Séance 2 (11 octobre) :	7
Séance 3 (18 octobre) :	8
Évaluation	8
Exercices durant les séances : 20%	8
Exercices hebdomadaires : 30%	8
Travail final : 50%	8
Références et ressources	9
Intégrité	9
Crédits	9
License	9
 I session 1	 10
1 Introduction	11
1.1 Lecture du plan de cours	11
1.2 Déroulement de l'atelier	11
1.3 Tour de table	11
1.4 En ce qui me concerne	11
1.5 Me contacter	12
 2 Les formats : théorie	 13
2.1 Les formats : un peu d'étymologie	13
2.2 Vers une définition du format	13
2.3 Format et édition savante	13

2.4	Format et extension	14
2.5	Formats libres, formats propriétaires	14
2.6	Le format n'est pas neutre !	14
2.7	Critères d'un format	14
2.7.1	Pertinence	15
2.7.2	Économie	15
2.7.3	Efficacité	15
2.7.4	Sécurité	15
2.8	Format, logiciel, extension	15
2.9	Structuration de l'information	16
2.10	Formats de balisage	16
2.10.1	Texte brut, texte enrichi	16
2.10.2	WYSIWYM vs WYSIWYG	16
2.10.3	WYSIWYM	17
2.11	Mettre en forme \neq structurer	17
2.12	Markdown	17
2.13	LaTeX	18
2.14	CSV	18
3	Exercice markdown	19
3.1	Markdown : langage de balisage léger	19
3.2	Rendu pour le 11 octobre	19
4	Bonnes pratiques d'édition	20
4.1	Les fichiers dans votre ordinateur	20
4.1.1	Organisation des fichiers	20
4.1.2	Nommage de fichier	20
5	Le plein texte	22
5.1	Plein / Plain	22
5.2	Un exemple : The Plain Text Project	22
6	Stylo	23
6.1	Chaînes modulaires	23
6.2	Un projet de recherche & un éditeur de texte	23
6.3	Stylo en mode avancé	23
6.4	Rendu pour le 18 octobre	23
7	Zotero	24
7.1	Zotero : introduction	24
7.2	Better Bibtex	24

II	session 2	25
8	Le <i>single source publishing</i>	26
8.1	Kézako	26
8.2	Intérêts	26
8.3	Limites	26
8.4	Quelques exemples	26
9	Utiliser son terminal	27
9.1	Terminal	27
9.2	À quoi ça sert	27
9.3	Quelques commandes	27
9.4	Rendu pour le 18 octobre	27
10	Convertisseurs : le cas de <i>pandoc</i>	28
III	session 3	29
11	Des caractères à l'écran – encodage numérique	30
11.1	Encodage numérique	30
11.2	Mémoire	30
11.3	0 ou 1 ?	30
11.4	Normes d'encodage	30
11.5	Encodage des données textuelles	30
11.5.1	ASCII	30
11.5.2	ISO 8859-1	30
11.5.3	UTF-8	30
11.6	Mauvais encodage/décodage	31
11.7	Une norme ?	31
11.8	Enjeux	32
12	Quarto	33
13	Git et GitHub	34
13.1	Versionnement	34
13.2	Git vs Github	34
14	YAML	35
14.1	À venir	35
	Références	36

Bienvenue

Bienvenue sur le support du cours **HNU6051 – Humanités numériques : formats d’écriture**. Celui-ci est destiné aux étudiantes et étudiants qui suivent l’atelier à l’automne 2024. Les séances se tiendront les vendredi 4, 11, 18 octobre, de 9h30 à 11h30 puis de 12h30 à 15h30 au Pavillon Lionel-Groulx, salle C-3115.

L’objectif de cet atelier est de découvrir les différents formats d’écriture et de prendre en main le fonctionnement de différentes chaînes éditoriales de production de contenu.

Aujourd’hui de nombreux outils ou applications de rédaction, d’édition et de publication de contenus textuels sont accessibles au grand public sur le Web, mais sans nécessairement prendre en compte les besoins et les pratiques d’écriture des communautés de recherche. La question des formats se place au cœur de la structuration du texte numérique et il est important, pour qui veut produire des connaissances organisées, pérennes et interopérables, de comprendre les enjeux culturels, techniques et politiques des formats.

Ces trois séances entendent donner aux étudiantes et étudiants une connaissance théorique et pratique sur les différents formats d’écriture, leur donner les clés pour mettre en place de bonnes pratiques éditoriales et de les former à quelques outils propices à la rédaction scientifique.

Support de cours

Le présent support est destiné aux étudiantes et étudiants de l’Université de Montréal suivant l’atelier [HNU6051, Humanités numériques : formats d’écriture](#), au trimestre d’Automne 2024.

La totalité du contenu de ce site est sous licence CC BY-NC-SA et est la propriété intellectuelle de Mathilde Verstraete.

Crédits

Le contenu de ce cours doit beaucoup aux préparations réalisées par [Antoine Fauchié](#), [Margot Mellet](#), [Alix Chagué](#) ainsi qu’aux séances de *Débogue tes humanités*.

Le support du cours a été créé avec [quarto](#).

Plan de cours

Informations pratiques

- Sigle du cours : HNU6051
- Intitulé du cours : Humanités numériques : formats d'écriture
- Chargée de cours : [Mathilde Verstraete](#)
- Horaire : trois vendredi : 4, 11, 18 octobre 2024, de 9h30 à 11h30 puis de 12h30 à 15h30
- Lieu : Pavillon Lionel-Groulx, salle C-3115.

Description et objectifs de l'atelier

La question des formats se place au cœur de la structuration du texte numérique et il est important, pour qui veut produire des connaissances organisées, pérennes et interopérables, de comprendre les enjeux culturels, techniques et politiques des formats. Aujourd'hui de nombreux outils ou applications de rédaction, d'édition et de publication de contenus textuels sont accessibles au grand public sur le Web. Ces outils ne prennent pas nécessairement en compte les besoins et les pratiques d'écriture des communautés de recherche.

Ces trois séances entendent donner aux étudiantes et étudiants une connaissance théorique et pratique sur les différents formats d'écriture, leur donner les clés pour mettre en place de bonnes pratiques éditoriales et les former à quelques outils propices à la rédaction scientifique.

L'atelier n'a pas seulement pour but de former à des outils spécifiques des Humanités numériques mais vise à offrir aux étudiant·e·s les connaissances théoriques et pratiques fondamentales pour leur permettre d'explorer les potentialités du numérique, de constituer une réflexion critique sur les outils numériques et de développer leurs propres pratiques d'écriture et de publication.

Support de cours

Le support de cours est multiformat. Un site Web est à votre disposition à l'adresse suivante : <https://mathildevrst.github.io/hnu6051-a24/>. Vous pourrez y télécharger le PDF et afficher les slides de chaque séance.

Organisation des séances

Nous alternerons entre pratique théorie et pratique. Il est conseillé de venir en cours avec son ordinateur personnel.

! Important

Le programme des séances est sujet à de légers changements en fonction du niveau et de l'avancement du groupe.

Séance 1 (4 octobre) :

Horaire	Matière	Type
09h30-10h00	Présentation & Introduction de l'atelier	–
10h00-11h00	Édition savante, formats de fichiers	Théorie
10h50-11h10	Bonnes pratiques de l'édition	Théorie
11h10-11h30	The Plain Text Project	Théorie et exercice
11h30-12h30	Pause	–
12h30-13h00	Markdown	Théorie et exercice
13h00-14h15	Chaînes éditoriales, Stylo	Théorie et pratique
14h15-15h15	Références structurées, Zotero	Théorie et exercice
15h15-15h30	Présentation du rendu pour la séance 2 (stylo basique)	Rendu

Séance 2 (11 octobre) :

Horaire	Matière	Type
09h30-10h00	Retour sur le rendu	Rendu
10h00-11h30	Single Source Publishing	Théorie
11h30-12h30	Pause	–
12h30-13h30	Utilisation du terminal, installation de Codium	Pratique
13h30-14h30	Zotero et Bibtex	Pratique
14h30-15h15	Stylo en mode avancé	Exercice
15h15-15h30	Présentation du rendu pour la séance 3 (stylo avancé ; recherches terminal)	Rendu

Séance 3 (18 octobre) :

Horaire	Matière	Type
09h30-10h00	Retour sur le rendu	Rendu
10h00-11h30	Les caractères à l'écran	Théorie
11h30-12h30	Pause	—
12h30-13h30	Quarto	Pratique
13h30-14h30	Git et Github	Théorie
14h30-15h15	Yaml	Théorie et exercices
15h15-15h30	Présentation du rendu final (création multi-format avec quarto)	Rendu

Évaluation**Exercices durant les séances : 20%**

Durant les séances de l'atelier, plusieurs exercices seront demandés à l'étudiant·e pour valider au fur et à mesure les acquis techniques. Cette évaluation se fera sur un mode continu.

Exercices hebdomadaires : 30%

Chaque semaine, en préparation de la séance prochaine, un exercice sera soumis à l'étudiant·e et sera discuté par la suite en classe :

- pour le 11 octobre : [Markdown](#) (10%)
- pour le 18 octobre : [Stylo en mode avancé](#) (10%); [Apprenez-moi quelque chose quant au terminal](#) (10%)

Travail final : 50%

Le travail final consiste en la création d'un *book* multiformat avec le système de publication [Quarto](#). Le contenu du travail est libre. Quelques critères sont à remplir :

- Le livre doit être présenté en deux formats au moins ;
- Le livre doit comporter au moins 4 *chapitres* ;
- Le thème doit avoir été modifié ;
- Le fichier `_quarto.yml` doit avoir été personnalisé ;
- Une publication accessible en ligne (via *github pages*) sera particulièrement appréciée ;
- Date de remise : vendredi 8 novembre.

Références et ressources

Une liste de références et de ressources variées sera mise à jour tout au long du cours. Elle est disponible [en cliquant ici](#).

Intégrité

L'Université de Montréal s'est dotée de deux règlements disciplinaires sur la fraude et le plagiat. L'une s'adresse aux étudiant·e·s de 1^{er} cycle, la seconde aux étudiant·e·s des cycles supérieurs : <https://integrite.umontreal.ca/boite-a-outils/les-reglements>.

Crédits

Le contenu de ce cours doit beaucoup aux préparations réalisées par [Antoine Fauché](#), [Margot Mellet](#), [Alix Chagué](#) ainsi qu'aux séances de *Débogue tes humanités*.

Le support du cours a été créé avec [quarto](#).

License

Tous les contenus de ce site sont sous licence [CC BY-NC-SA](#) : Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions.

Première partie
session 1

1 Introduction

1.1 Lecture du **plan de cours**

1.2 Déroulement de l'atelier

- Un peu de théorie
- Beaucoup de pratique : on met les mains à la pâte, on teste des outils, on produit, on fabrique !

Vous n'êtes pas un groupe homogène donc vous n'irez pas toutes et tous à la même vitesse, pas d'inquiétude !

1.3 Tour de table

Quelques informations pour savoir qui vous êtes, quel est l'objectif de cet atelier et comment les séances vont être organisées.

5 questions :

- quel est votre nom ?
- où étiez-vous l'année dernière ?
- quel cursus suivez-vous cette année ?
- pourquoi suivez-vous cet atelier ?
- qu'est-ce qui vous vient instinctivement à l'esprit quand je vous dis format ?

1.4 En ce qui me concerne

En ce qui me concerne, je m'appelle Mathilde Verstrate. Je suis doctorante au Département de littératures et de langues du monde à l'Université de Montréal, dans l'option humanités numériques. J'ai une formation en lettres classiques – que j'ai suivie en Belgique, à l'Université Catholique de Louvain. Je suis passée du côté obscur de la force en rejoignant la Chaire de recherche du Canada sur les écritures numériques (dirigée par Marcello Vitali-Rosati). C'est

à partir de ce moment là que j'ai commencé à m'intéresser et à me spécialiser dans les humanités numériques. J'étais donc grande débutante il y a peu, mais vous verrez, la courbe d'apprentissage n'est pas trop abrupte. C'est néanmoins la première fois que j'enseigne ce cours – j'enseignais un cours de latin les années précédentes.

1.5 Me contacter

N'hésitez pas à me contacter. Veillez à vous y prendre à l'avance si vous avez des questions concernant les rendus.

mathilde.verstraete@umontreal.ca

2 Les formats : théorie

2.1 Les formats : un peu d'étymologie

- Format, fromage ? ;
- Dimensions d'une feuille de papier, d'un feuillet ;
- Du latin *formare* : donner une forme, former, conformer ; arranger, organiser, régler ; façonner
- Extension à de nombreux domaines : imprimerie, photographie, cinéma, **informatique**

2.2 Vers une définition du format

- ce qui donne une forme et un ensemble de caractéristiques à un ensemble d'information (image, écriture, son, etc.) ;
- la dimension (format A4 ou *letter* par ex.) qui va impliquer un agencement ;
- une manière de représenter des données (pdf, docx) qui impose une structure

Structure définie de données contenues dans un fichier, établie selon des règles qui régissent notamment leur sauvegarde, leur affichage et leur manipulation par un logiciel compatible. ([OQLF, 2022](#)).

2.3 Format et édition savante

- Production de documents scientifiques qui ont des caractéristiques exigées par les normes de l'édition savante :
 - appareil critique accompagnant le texte (notes de bas de page, bibliographie, index, etc.) ;
 - citabilité (métadonnée → identification) ;
 - structuration (normalement) rigoureuse (balisage sémantique, non ambigu) ;
 - pérennité
- Les formats utilisés doivent répondre à ces exigences, permettant notamment d'explicit-ter la fonction de chaque élément du document (titre, citation, métadonnées, références bibliographiques,...)

2.4 Format et extension

À chaque format de fichier est associée une extension de nom de fichier qui permet de l'identifier. [Cette] extension de nom de fichier indique au système d'exploitation quel logiciel d'application peut ouvrir le fichier et quelle icône afficher. (OQLF, 2022)

L'extension d'un fichier texte indique à l'utilisateur et à l'ordinateur quel est son format, c'est-à-dire :

- comment sont organisées les informations relatives au fichiers et les métadonnées
- avec quel logiciel l'ouvrir
- s'il contient ou non des informations de mise en page qu'il faut interpréter

2.5 Formats libres, formats propriétaires

Formats libres	Formats propriétaires
ouverts	fermés (parfois ouverts)
spécifications techniques publiques	spécifications techniques contrôlées par une entité privée
pas de restriction d'accès	usage limité
définis par les pratiques	conditionnés par les pratiques
exigences de recherche	exigences commerciales
.txt ; .odt ; .tex ; .html ; .svg ; .png ; .epub ; .zip	.pdf ; .docx ; .pptx ; .xlsx ; .ai ; .jpeg ; .psd ; .mp3

2.6 Le format n'est pas neutre !

Le format est politique en ce qu'il porte un modèle de l'information.

2.7 Critères d'un format

Le format répond à plusieurs critères : pertinence, économie, efficacité, sécurité.

2.7.1 Pertinence

Certains formats sont plus pertinents que d'autres, dépendant de la discipline, de l'utilisation, de la langue, du contexte,...

Ex. : .pdf a été conçu spécifiquement pour la mise en page de documents prévus pour l'impression

2.7.2 Économie

Quel modèle soutient le format ? Quel poids du document ? Pour quelle possibilité de stockage ?

Ex. : un texte à envoyer rapidement (une note) doit être relativement léger ; pour un texte plus important, on peut prioriser des fonctionnalités plus complexes, plus permissives mais qui alourdissent le fichier.

2.7.3 Efficacité

Quelle temporalité pour l'écriture et la finalisation du document ?

2.7.4 Sécurité

Le contenu est-il garanti ? Quelles protection pour les données ? Quelle pérennité pour le document ?

2.8 Format, logiciel, extension

- un format n'est pas un logiciel : parler du format Word, c'est faux ;
- un format n'est pas une extension : changer l'extension ne change pas le format.

Le format va structurer une information, va lui donner une forme, va délimiter son utilisabilité, les manipulations qui vont être possibles ou non.

2.9 Structuration de l'information

Nous devons faire la distinction entre :

- le corps du document,
- les métadonnées,
- le contenu additionnel, comme la bibliographie.

2.10 Formats de balisage

- un format de balisage : des balises pour structurer et qualifier le texte (.md, .html, .xml) ;
- un format de balisage léger : compréhensible par les humains ;
- un format transformable : format source facilement interprétable pour obtenir un format de sortie riche
- un format pérenne : encore utilisable dans 10 ans (car porté par une communauté et des standards)

2.10.1 Texte brut, texte enrichi

Les informations de mise en forme sont des annotations qui sont formées selon plusieurs standards ou normes.

- texte brut (ou *plain text*) : texte dépourvu d'information de mise en forme, qui se limite à l'affichage de la traduction des données binaires en caractères alphanumériques et symboles ;
- texte enrichi (ou *rich text*) : texte qui comporte des éléments de mise en forme, qu'ils soient ou non affichés dans l'interface graphique ; Un fichier de texte peut contenir des indications de mise en forme.

Les normes et standards encadrant la manière de former les annotations sont étroitement liées au format du document et au standard qu'il suit.

2.10.2 WYSIWYM vs WYSIWYG

WYSIWYM	WYSIWYG
<i>What You See Is What You Mean</i> éditeur de texte qui présente un texte enrichi avec sa mise en forme, de manière simultanée	<i>What You See Is What You Get</i> éditeur de texte qui présente le texte et ses informations de mise en forme dans les interpréter

2.10.3 WYSIWYM

- dans un éditeur WYSIWYM : moindre confusion entre mise en forme et indication de la structure logique
- la mise en forme apparaît après parsing du document (création d'un PDF ou ouverture dans un navigateur)
- le paramétrage de la mise en forme peut être géré dans un fichier externe (ex : CSS pour HTML)



2.11 Mettre en forme =/= structurer

Mise en forme	Structure logique
Du texte en gras	Un titre
Du texte en italique	Un titre, un mot étranger
Du texte entre guillemets	Une citation

2.12 Markdown

- Inventé en 2004 par John Gruber et Aaron Swartz ;
- Markdown est un langage de balisage léger qui permet d'encoder facilement la structure et certains éléments de mise en forme du texte sans gêner sa lecture ;
- Il permet de créer des fichiers HTML ou PDF et est beaucoup utilisé sur le Web, notamment sur des plateformes comme Github ;
- Même non compilé, un texte Markdown est facile à lire car les annotations sont légères et limitées ;
- On peut utiliser Markdown ou un système similaire dans Whatsapp et Messenger.

(Plus d'informations : Simpkin 2015)

Exercice : Faire ce tuto (en classe) : <https://www.arthurperret.fr/tutomd/>

2.13 LaTeX

- LaTeX est un langage et un système de composition de documents, bien plus complet que Markdown ;
- Utilisé dans le domaine des sciences car on peut facilement écrire des formules mathématiques :
- Les balises sémantiques de LaTeX sont plus “invasives”. Exemple : `\chapter{Titre de chapitre}` ;
- Pour voir la mise en forme, on doit *compiler* le fichier .tex. On obtient généralement un PDF ;
- LaTeX ne sert pas qu'à annoter la structuration logique : il prend en charge la mise en forme du document au sens large.

2.14 CSV

CSV (Comma Separated Values) - données tabulaires

- 1 ligne de texte = 1 ligne de tableau ;
- 1 virgule = 1 nouvelle cellule de tableau ;
- un fichier csv peut être ouvert avec Excel.

Id_Monographie,Id_Enquête,Serie_Tome,Prénom,Nom,Place dans la famille 001a,001aE1,1-1 (1857),Jean,M,**Père 001a,001aE2,1-1 (1857),Marie,R**,Mère 001a,001aE3,1-1 (1857),Joseph,M**,Fils,

3 Exercice markdown

3.1 Markdown : langage de balisage léger

Markdown	Rendu graphique	Balisage HTML
Un texte <code>_en italique_</code>	Un texte <i>en italique</i>	Un texte <code>en italique</code>
<code>### Titre de niveau 3</code>	Titre de niveau 3	<code><h3>Titre de niveau 3</h3></code>
Un <code>[lien] (https://www.unlien.org)</code>	Un lien	Un <code>lien</code>
Un texte <code>**en gras**</code>	Un texte en gras	Un texte <code>en gras</code>

3.2 Rendu pour le 11 octobre

- Texte : [pont_mirabeau.txt](#)
- Image : [pont_mirabeau.png](#)
- Éditeur en ligne : [Dillinger](#)

Faire un encodage en Markdown du poème :

- mettre le titre en niveau 1
- faire de chaque première ligne de strophe un titre de niveau 3
- mettre en italique le paratexte
- mettre les noms propres en gras
- sauver le document en `.md`
- le déposer sur studium

4 Bonnes pratiques d'édition

4.1 Les fichiers dans votre ordinateur

4.1.1 Organisation des fichiers

- votre ordinateur a son propre système d'organisation ;
- il faut comprendre les systèmes de chemin ;
- tout est arborescence

Il faut d'abord comprendre comment est organisé votre ordinateur et où se trouvent vos fichiers. Votre disque dur peut être compris comme une grande boîte dans laquelle se trouvent d'autres boîtes (dossiers). La boîte principale est ce qu'on appelle « racine ». Dans un ordinateur Unix (Mac ou Linux) la racine est /. Dans un ordinateur Windows, la racine est C :/. Dans la racine vous trouvez plusieurs « boîtes ».

L'emplacement du Dossier correspond à ce qu'on appelle son « chemin » (path).

Exemple d'un dossier « Vacances » :

- Unix : /Document Personnel/Images/Vacances
- Windows : C :/Dossier Personnel/Images/Vacances

Il est important que vous connaissiez la structure de l'arborescence de votre ordinateur car c'est ainsi qu'on accède aux fichiers.

4.1.2 Nommage de fichier

Quel nom donner à un fichier ? Quelques règles :

- les noms de fichier ne doivent pas contenir des caractères spéciaux, ni des accents, ni des espaces ;
- les noms des fichiers doivent être significatifs ;
- les noms des fichiers ne doivent pas dépasser 80 caractères ;
- faites attention aux extensions de fichiers !

Exemple : 2024-10-04-edn6051-bonnes-pratiques.md

Cette question pas si banale que ça. Il existe des règles qui sont fondamentales pour :

- classer un fichier,
- le retrouver facilement,
- partager ses fichiers avec d'autres personnes,
- éviter des problèmes de comptabilité d'un système à un autre.

5 Le plein texte

5.1 Plein / Plain

5.2 Un exemple : [The Plain Text Project](#)

6 Stylo

6.1 Chaînes modulaires

6.2 Un projet de recherche & un éditeur de texte

6.3 Stylo en mode avancé

6.4 Rendu pour le 18 octobre

Création d'un article à partir de l'article « La traduction de documents XML dans les mémoires de traduction » (https://www.persee.fr/doc/equiv_0751-9532_2005_num_32_1_890) de Pascaline Merten :

Méthode :

- parcourir le document et relevé les différentes tâches à réaliser (structuration, métadonnées) ;
- commencer par intégrer le texte en copiant-collant la « version texte » – à retrouver sur la page Persée –, puis en le structurant en vous référant à la version graphique : niveaux de titres, italiques, notes de bas de page, références bibliographiques, images ;
- ajouter au moins quelques références bibliographiques ;
- ajouter toutes les métadonnées dont vous disposez via le mode « Editor Mode ».

Une fois le travail terminé, partagez votre article avec moi (mathilde.verstraete@umontreal.ca) et [prévenez-moi par courriel](#).

7 Zotero

7.1 Zotero : introduction

7.2 Better Bibtex

Deuxième partie

session 2

8 Le *single source publishing*

8.1 Kézako

8.2 Intérêts

8.3 Limites

Avez-vous remarquer les émojis dans mon plan de cours ? Sont-ils dans la sortie PDF ?

8.4 Quelques exemples

9 Utiliser son terminal

9.1 Terminal

9.2 À quoi ça sert

9.3 Quelques commandes

9.4 Rendu pour le 18 octobre

Apprenez-moi quelque chose à propos du terminal! Cela peut concerner une commande que vous trouvez utile, une astuce pour gagner du temps, ou une fonctionnalité du terminal que peu de gens connaissent. Nous commencerons le cours du 18 par vos trouvailles : une simple explication suffit, une démonstration est bienvenue.

10 Convertisseurs : le cas de *pandoc*

Troisième partie

session 3

11 Des caractères à l'écran – encodage numérique

11.1 Encodage numérique

11.2 Mémoire

11.3 0 ou 1 ?

Tout document conservé sur un ordinateur est constitué d'une série de 0 et de 1.

Ce sont les normes d'encodage et les formats qui permettent d'interpréter le sens de ces 0 et 1.

11.4 Normes d'encodage

11.5 Encodage des données textuelles

11.5.1 ASCII

11.5.2 ISO 8859-1

11.5.3 UTF-8

- [Universal Character Set Transformation Format](#) + base d'encodage (8, 16, 32, ...)
- En fonction de la base d'encodage, un caractère peut être représenté par 1, 2, 3 ou 4 octets.
- UTF est rétrocompatible avec ASCII, mais pas avec ISO 8859-1. Il s'inscrit dans le cadre établi par Unicode qui vise à faciliter l'échange de textes numériques en assurant leur compatibilité.
- Unicode définit plus de 137 000 caractères, issus de multiples alphabets, et prévoit aussi des symboles comme les emojis.

— Exemple : = 11110000 10011111 10100100 10010011

11.6 Mauvais encodage/décodage

Si on n'utilise pas la bonne clef d'encodage ou de décodage d'un texte, on risque d'obtenir un texte partiellement ou entièrement illisible.

UTF-8 ouvert avec 8859-1 (Latin 1)

La Joconde, ou Portrait de Mona Lisa, est un tableau de l'artiste Léonard de Vinci, réalisé entre 1503 et 1506 ou entre 1513 et 1516^{1,2}, et peut-être jusqu'à 1519 (l'artiste étant mort cette année-là, le 2 mai)³, qui représente un portrait mi-corps, probablement celui de la Florentine Lisa Gherardini, épouse de Francesco del Giocondo. Acquise par François Ier, cette peinture à l'huile sur panneau de bois de peuplier de 77 × 53 cm est exposée au musée du Louvre à Paris. La Joconde est l'un des rares tableaux attribués de façon certaine à Léonard de Vinci.

UTF-8 ouvert avec ISO 8859-5 (Cyrillic)

La Joconde, ou Portrait de Mona Lisa, est un tableau de l'artiste L onard de Vinci, r alis entre 1503 et 1506 ou entre 1513 et 1516^{1,2}, et peut- tre jusqu' 1519 (l'artiste tant mort cette ann e-l , le 2 mai)³, qui repr sente un portrait mi-corps, probablement celui de la Florentine Lisa Gherardini, pouse de Francesco del Giocondo. Acquise par Fran ois Ier, cette peinture l'huile sur panneau de bois de peuplier de 77 53 cm est expos e au mus e du Louvre Paris. La Joconde est l'un des rares tableaux attribu s de fa on certaine L onard de Vinci.

11.7 Une norme ?

UTF s'impose désormais comme la norme

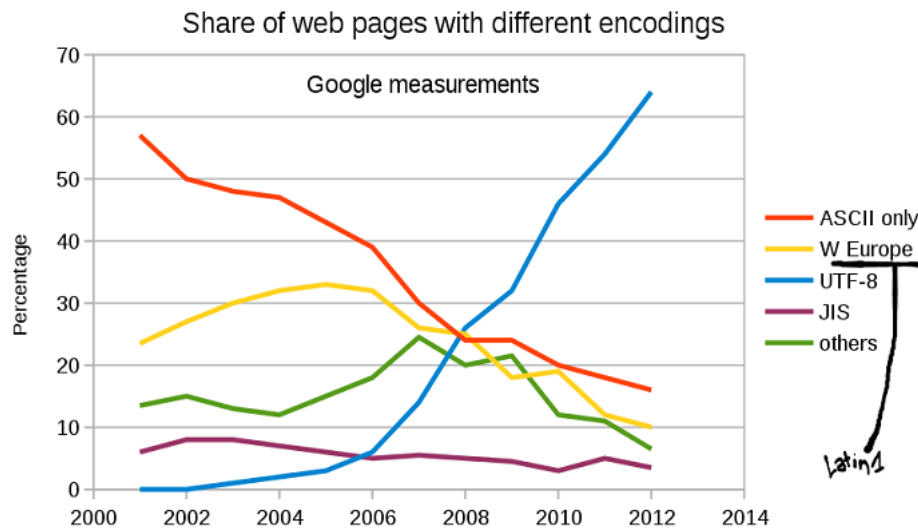


FIGURE 11.1 – diagramme montrant qu'UTF s'impose sur les autres normes depuis 2006

D'après W3Techs, 97,8% des pages web qui déclarent un encodage suivent UTF-8 aujourd'hui.

(source du graphique : Wikimedia)

11.8 Enjeux

12 Quarto

13 Git et GitHub

13.1 Versionnement

13.2 Git vs Github

14 YAML

14.1 À venir

Références

- Boulétreau, Viviane, and Benoît Habert. 2014. “Les formats.” In *Pratiques de l’édition numérique*. Les Ateliers de [sens public].
- Merten, Pascaline. 2005. “La traduction de documents XML dans les mémoires de traduction.” *Equivalences* 32 (1) : 63–78. <https://doi.org/10.3406/equiv.2005.890>.
- Simpkin, Sarah. 2015. “Débuter avec Markdown.” *Programming Historian*, November.