

HNU6051-A24

Humanités numériques : formats d'écriture

Mathilde Verstraete

Automne 2024

Contenu

Bienvenue	4
Support de cours	4
Crédits	4
Plan de cours	5
Informations pratiques	5
Description de l'atelier	5
Objectifs de l'atelier	5
Support de cours	6
Organisation des séances	6
Séance 1 (4 octobre) :	6
Séance 2 (11 octobre) :	6
Séance 3 (18 octobre) :	7
Évaluation	7
Exercices durant les séances : 20%	7
Exercices hebdomadaires : 30%	7
Travail final : 50%	8
Références et ressources	8
Intégrité	8
Crédits	8
License	8
1 Introduction	9
1.1 Lecture du plan de cours	9
1.2 Déroulement de l'atelier	9
1.3 Tour de table	9
1.4 En ce qui me concerne	9
1.5 Me contacter	10
2 Les formats : théorie	11
2.1 Les formats : un peu d'étymologie	11
2.2 Structure d'un document numérique	11
2.3 Format : définition	11
2.4 Format et extension	11
2.5 Formats ouverts, formats fermés	11

2.6	Formats propriétaires, formats libres	11
2.7	Formats sources, formats de sortie	11
2.8	Structuration de l'information	11
2.9	Formats de balisage	11
2.9.1	Texte brut, texte enrichi	11
2.9.2	WYSIWYM	11
2.10	Mettre en forme \neq structurer	11
2.11	Structure logique vs apparence	11
2.12	Markdown	12
2.12.1	Exercice markdown	12
2.13	LaTeX	12
2.14	CSV	12
3	Des caractères à l'écran – encodage numérique	14
3.1	Encodage numérique	14
3.2	Mémoire	14
3.3	0 ou 1 ?	14
3.4	Normes d'encodage	14
3.5	Encodage des données textuelles	14
3.5.1	ASCII	14
3.5.2	ISO 8859-1	14
3.5.3	UTF-8	14
3.6	Mauvais encodage/décodage	15
3.7	Une norme ?	15
3.8	Enjeux	16

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 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.

Objectifs de l'atelier

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.

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.

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, bonnes pratiques	Théorie
11h00-11h30	Markdown	Théorie et exercice
11h30-12h30	Pause	–
12h30-13h00	The Plain Text Project	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

Horaire	Matière	Type
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 : Stylo en mode basique (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

Voici une liste de ressources utiles :

Simpkin, Sarah. 2015. “Débuter avec Markdown.” *Programming Historian*, November.

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 Fauchié, 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.

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

2.2 Structure d'un document numérique

2.3 Format : définition

2.4 Format et extension

2.5 Formats ouverts, formats fermés

2.6 Formats propriétaires, formats libres

2.7 Formats sources, formats de sortie

2.8 Structuration de l'information

2.9 Formats de balisage

2.9.1 Texte brut, texte enrichi

2.9.2 WYSIWYM

2.10 Mettre en forme \neq structurer

2.11 Structure logique vs apparence

*L'ouvrage *Pratiques de l'édition numérique* est a priori.*

- tous les mots en italiques ne sont pas des mots étrangers, mais tous les mots étrangers sont en italiques
- titres d'ouvrages cités en italiques (ou pas) et mots étrangers en italiques (ou pas)
- on peut extraire rapidement tous les mots étrangers sans récolter aussi les titres d'ouvrages cités

2.12 Markdown

Voir (Simpkin 2015)

2.12.1 Exercice markdown

- 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

2.13 LaTeX

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_Enqueté	Serie_Tome	Prénom	Nom	Place dans la famille	Lieu de nais
001a	001aE1	1-1 (1857)	Jean	M**	Père	T** (Aube)
001a	001aE2	1-1 (1857)	Marie	R**	Mère	L** (Meurthe)
001a	001aE3	1-1 (1857)	Joseph	M**	Fils	

Id_Monographie,Id_Enqueté,Serie_Tome,Prénom,Nom,Place dans la famille,Lieu de naissance 001a,001aE1,1-1 (1857),Jean,M,**Père,T** (Aube) 001a,001aE2,1-1 (1857),Marie,R,**Mère,L** (Meurthe) 001a,001aE3,1-1 (1857),Joseph,M**,Fils,,

2.15

3 Des caractères à l'écran – encodage numérique

3.1 Encodage numérique

3.2 Mémoire

3.3 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.

3.4 Normes d'encodage

3.5 Encodage des données textuelles

3.5.1 ASCII

3.5.2 ISO 8859-1

3.5.3 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

3.6 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.

3.7 Une norme ?

UTF s'impose désormais comme la norme

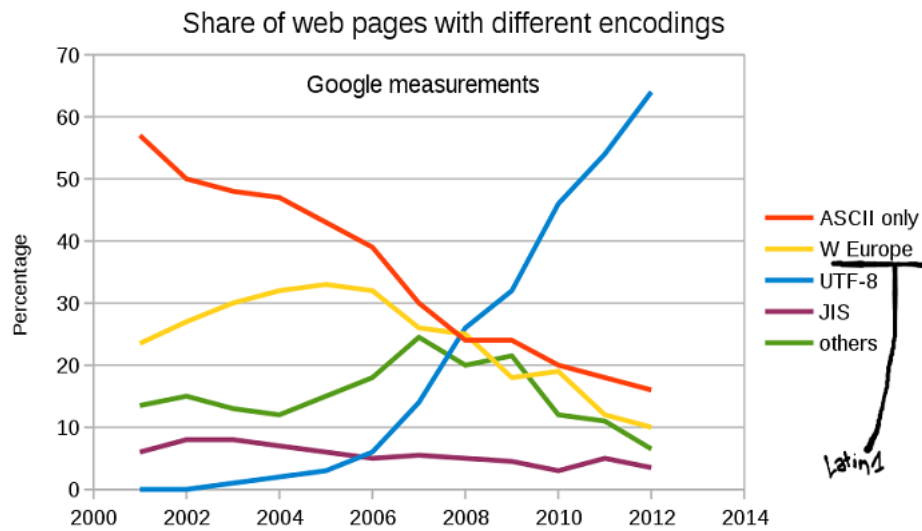


FIGURE 3.1 – diagramme montrant qu'UTF s'impose sur les autres normes depuis 2006

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

(source du graphique : Wikimedia)

3.8 Enjeux