HNU2000-A25

Humanités Numériques : technologies

Mathilde Verstraete

Automne 2025

Plan du cours

• Sigle du cours : HNU2000

• Titre du cours : Humanités numériques : technologies

• *Trimestre*: Automne 2025

• Titulaire du cours : Mathilde Verstraete

• Coordonnées: mathilde.verstraete@umontreal.ca

• Horaire: Mardi 08h30-11h30 (du 02/09 au 09/12/2025)

• Lieu: Cf. votre centre étudiant

Description du cours

Espace de formation pratique basée sur des compétences pour les méthodes de base dans les humanités numériques, y compris l'exploration de textes, l'analyse de données, l'analyse du web et les systèmes d'information géographique.

Objectifs et contenu du cours

Le numérique habite l'ensemble de nos vies et touche aussi, et surtout, à nos activités purement « humanistes », ou même « humaines ».

(Sinatra & Vitali-Rosati, 2014)

Les disciplines humanistes au sens large s'appuient de plus en plus sur des outils informatiques pour explorer, traiter, analyser, diffuser leurs objets d'étude. Cette irruption des outils issus des sciences dites dures dans les sciences humaines a donné naissance à un nouveau champ interdisciplinaire, celui dit des *humanités numériques*.

L'objectif principal du cours *HNU2000 – Humanités numériques : technologies* est de familiariser les étudiant-e-s aux principaux outils, technologies, logiciels, utilisés dans ce champ. Cette familiarisation passera par la manipulation concrète d'outils numériques et par la réalisation de projets pratiques.

Parmi les thématiques abordées durant le cours:

- La connaissance de son ordinateur ;
- L'encodage et la structure de fichiers numériques ;

- Les formats d'écriture ;
- Le versionnement de ses données ;
- Le traitement et la gestion des références bibliographiques ;
- La fouille et l'analyse de textes ;
- La reconnaissance automatique d'écritures manuscrites ;
- etc.

Les « humanités numériques » sont à la fois une méthode scientifique, un programme de recherche ou une approche pluridisciplinaire : elles offrent de nouvelles perspectives pour appréhender, lire et comprendre le monde qui nous entoure. Depuis l'avènement du numérique, la compréhension des écosystèmes technologiques devient une nécessité dans le champ des sciences humaines. Le cours HNU2000 « Humanités numériques : technologies » est une opportunité pour explorer de façon originale les outils théoriques et pratiques utilisés dans les humanités numériques : lire, écrire, chercher, explorer, visualiser, analyser, publier, etc. Ce cours est un espace de formation pratique et de découverte des méthodes de base en humanités numériques, avec une orientation vers les démarches d'écriture, d'édition et de publication.

Ce cours est une initiation aux technologies utilisées dans les humanités numériques, les étudiant-e-s devront comprendre, explorer, manipuler et expérimenter des outils, des logiciels, des méthodes et des programmes informatiques. Le cours HNU1000 « Humanités numériques : théories » est un compagnon adéquat pour ce cours pratique.

À l'issue du cours les étudiant·e·s seront en mesure de comprendre les enjeux technologiques des humanités numériques, de réutiliser des concepts liés aux humanités numériques, d'appréhender des méthodes utilisées dans les humanités numériques et d'utiliser une multitude d'outils/applications.

L'outillage est souvent un aspect mal considéré en humanités, pourtant c'est la condition même de la possibilité de nos recherches : nos outils de collecte, d'enregistrement, de traitement, de prise de notes, de rédaction et de publication sont ce par quoi tout commence.

Organisation du cours

Le cours alterna entre théorie et pratique. De manière générale, chaque cours sera l'occasion d'étudier un outil. À cette fin, il sera demandé à l'étudiant e de préparer la séance par une ou des lecture(s) et/ou des manipulations sur son ordinateur personnel (installation de logiciel, création de compte, etc.). Le cours sera l'occasion d'étudier – d'un point de vue pratique, mais aussi théorique – l'outil en question. Il pourra être demandé à l'étudiant e de terminer des manipulations après la séance de cours et pour le cours suivant.

Dans le calendrier, vous trouverez sous chaque séance ce qui sera abordé, les lectures et actions (téléchargement, création de compte, etc.) à faire avant la séance ainsi que des lectures suggérées.

! Important

Il est impératif que l'étudiant e possède un ordinateur et l'apporte en cours.

Modalités d'évaluation

Important

Rappel: 1h de cours = 1,5h de travail hors cours

- Participation ¹ & Implication dans les exercices et travaux non évalués : 25%;
- 28/10 Examen de mi-session : 25%;
- 28/10 Travail bonus : Apprenez-moi quelque chose ;
- 09/12 Travail final: 25%;
- 16/12 Présentation finale: 25%.

Examen de mi-session

- *Modalités* : examen sur table ;
- *Durée*: environ 2h;
- *Forme* :
 - Questions à choix multiple;
 - Questions ouvertes;
- Contenu : questions théoriques portant sur :
 - Le cours;
 - Les lectures obligatoires.

Travail bonus : Apprenez-moi quelque chose

- Ce travail n'est pas obligatoire :
 - Si vous ne le rendez pas → aucune pénalité;
 - Si vous le rendez → possibilité de gagner des points supplémentaires sur la note globale du cours.
- Consignes:
 - Longueur: entre 1.5 et 3 pages;
 - Sujet : description d'un outil, projet ou actualité de votre choix, en lien avec les humanités numériques;
 - Contraintes techniques : doit être rédigé avec Stylo ou LaTeX;
- Éléments évalués :

¹Par participation, j'entends la préparation, la concentration lors des séances, l'écoute active, l'implication lors des échanges.

Qualité de la présentation générale (orthographe, structure, mise en page, ajout de liens, bibliographie minimale);

- Clarté et pertinence de la description du projet choisi ;
- Force de l'argumentation qui doit donner envie d'en savoir plus ;
- Impact sur la note finale :
 - En fonction de la qualité du travail, votre note peut monter de deux niveaux (ex. : un B+ peut devenir un A- ou un A);
- Date de remise : 28 octobre (aucun délai ne sera accordé).

Examen final

Travail écrit

- Date de remise : au plus tard le lundi 9 décembre ;
 - Une pénalité de 10% sera appliquée par jour de retard ;
- Format attendu:
 - Analyse d'un outil présenté en classe ou d'un autre outil (avec approbation préalable de la chargée de cours);
 - Présentation soignée : orthographe, mise en page, ajout éventuel d'images, bibliographie, liens utiles;
 - PDF, HTML ou autre soyez créatifs (évitez les traitements de texte!);
 - Environ 5 à 7 pages (hors bibliographie);
- Contenu attendu:
 - Présentation de l'outil (objectifs, fonctionnement, contexte d'utilisation);
 - Analyse critique (points forts, limites, potentiel pour les humanités numériques);
 - Éventuelle mise en dialogue avec des articles évoquant l'outil en question ;
 - Mise en perspective par rapport aux thématiques du cours.

Présentation orale et discussion en classe

- Date: lundi 16 décembre;
- *Durée* : environ 10 minutes de présentation + 5 minutes de discussion ;
- Contenu attendu:
 - Présentation synthétique de l'outil analysé dans votre travail écrit ;
 - Mise en avant des aspects les plus intéressants pour susciter l'échange avec vos collègues;
 - Ouverture à la discussion : préparez une ou deux questions pour la classe.

Calendrier des séances et des lectures

Le cours de cette année sera axé sur le cycle de vie du texte savant en contexte numérique. Nous aborderons le texte source (séances 2-6), son analyse (8-10) et sa publication et diffusion (11-13).

Pour une description détaillée des séances, lectures obligatoires et conseillées, cf. le calendrier des séances.

Ressources utilisées dans le cadre de ce cours

- Le présent site/syllabus contient l'ensemble des informations dont vous aurez besoin pour ce cours.
- Lorsque les références sont disponibles en ligne le lien est indiqué sur le site du cours. Attention, certaines ressources en ligne nécessitent une connexion UdeM : soit depuis l'Université, soit en passant par le VPN.

Plagiat et fraude

« Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par une étudiante, un étudiant, de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen ou d'un travail faisant l'objet d'une évaluation ou dans toute autre circonstance, constituent une infraction. Vous trouverez à l'adresse suivante les différentes formes de fraude et de plagiat ainsi que les sanctions prévues par l'Université : https://integrite.umontreal.ca »

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 et à l'ouvrage *Vade-mecum informatique pour les lettres et sciences humaines* (Debouy, 2025).

Le support du cours a été créé avec Quarto.

License

Tous les contenus de ce site ou de ce document sont sous licence CC BY-NC-SA : Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions.

Calendrier des séances et des lectures

[Séance 1:02/09] Introduction au cours et aux Humanités Numériques

- Présentation de l'enseignante et des étudiant es;
- Présentation du cours : déroulement, courte présentation des séances, participation, évaluations ;
- Introduction aux Humanités Numériques.

Lectures supplémentaires

- Dacos & Mounier (2015)
- Dacos (2011)
- Burdick et al. (2012)
- Burnard (2012)
- Underwood (2018)

[Séance 2: 09/09] COMPRENDRE son environnement

- Prise en main de son ordinateur : fichiers, terminal, ligne de commande ;
- Bash :
- L'éditeur vs le traitement de texte.
- Lectures obligatoires
 - Sinatra & Vitali-Rosati (2014)
- l Outils à télécharger
 - https://vscodium.com/#install.
- Lectures supplémentaires
 - Schmidt (2016)

[Séance 3:16/09] CHERCHER: corpus et ressources numériques

Note

[17/09] Date limite pour la modification des choix de cours.

- Internet et le Web;
- Les navigateurs et moteurs de recherche;
- Une recherche efficace;
- Manipulation : Isidore.
- Lectures obligatoires
 - France Inter (s. d.) (podcast)
 - Huma-Num (2021)
- Outils à télécharger
 - Créez-vous un compte HumanID
- Lectures supplémentaires
 - Cardon (2013)
 - Larsonneur (2008), sur la question des navigateurs et moteurs de recherche, notamment
 - Ochigame & Ye (2021), sur les biais géographiques des résultats aux recherches sur Google

[Séance 4 : 23/09] ORGANISER & BIBLIOGRAPHIER : métadonnées et formats

- Les métadonnées ;
- Les différents formats textuels ;
- La bibliographie : définition, intérêts, mode d'emploi ;
- Manipulation : Zotero.
- Lectures obligatoires
 - Collectif (2018)
- Outils à télécharger
 - Téléchargez Zotero (Zotero & l'extension Zotero Connector);
 - Créez-vous un compte Zotero
- Lectures supplémentaires
 - Schrijver (2017)
 - Daniel (2022)

[30/09]

i Note

Congé universitaire – Journée nationale de la vérité et de la réconciliation.

[Séance 5:07/10] REPRÉSENTER le texte (I) : formats légers

- L'écriture numérique ;
- Les formats légers : .txt, .md, .html;
- Manipulation: Stylo.
- Lectures obligatoires
 - Vitali Rosati et al. (2020)
- Outils à télécharger
 - Créez-vous un compte/connectez-vous à Stylo via votre HumanID (cf. la Séance 3)
- Lectures supplémentaires
 - Vitali-Rosati (2018)

[Séance 6 : 14/10] REPRÉSENTER le texte (II) : formats riches et normés

Note

[07/11] Date limite pour l'abandon d'un cours.

- Les langages interprétés : l'exemple de LaTeX ;
- Introduction à la TEI;
- Manipulations:
 - LaTeX (via Overleaf);
 - Écrire en TEI-XML (via LeafWriter);
 - Pandoc si le temps le permet.
- Lectures obligatoires
 - Coombs et al. (1987)
- Outils à télécharger
 - Créez-vous un compte Overleaf
 - Téléchargez Pandoc : Nous le ferons ensemble en classe
- Lectures supplémentaires
 - Burnard (2015), lecture vivement conseillée
 - « Just What Is TeX? » (2023), lecture vivement conseillée
 - André (1990)

[21/10]

Note

Semaine de lecture - pas de cours.

[Séance 7:28/10] EXAMEN

- ! Important
 - Examen de mi-session;
 - Remise du travail facultatif.

[Séance 8:04/11] ANNOTER : des $\emph{marginalia}$ à l'annotation numérique

- Historique;
- Le Web: enjeux collaboratifs;Manipulation: Hypothes.is.
- Lectures obligatoires
 - Fredriksson & Sauret (2019)
- Outils à télécharger
 - Créez-vous un compte Hypothes.is
- Lectures supplémentaires
 - Jahjah (2017)

[Séance 9:11/11] ÉCRIRE & PUBLIER

- Le Single Source Publishing;
- La gestion de versions;
- Manipulation : Git et GitHub.
- ! Lectures obligatoires
 - Fauchié & Audin (2023)
- Outils à télécharger
 - Créez-vous un compte GitHub
- Lectures supplémentaires
 - Turbiau (2023)

[Séance 10:18/11] NETTOYER & EXPLOITER les données textuelles

Avertissement

- Conférence & Workshop, A. Chagué, « De l'OCR à l'HTR : panorama général, suivi d'un workshop introductif à eScriptorium » – À confirmer!
- La notion de *data* ;
- La numérisation des corpus : qui, quoi, comment ? ;
- Une petite histoire de l'OCR à l'HTR.
- Lectures obligatoires
 - Bermès (2023)

[Séance 11:25/11] ANALYSER : textométrie, stylométrie et TAL

- Humanités Numériques et Texte;
- Manipulations:
 - Ngram Viewer;
 - Voyant Tools.
- Lectures obligatoires
 - Reboul & Gefen (2019)
- Lectures supplémentaires
 - Bot (2012)

[Séance 12:02/12] VISUALISER : vers une autre lecture

- Les graphes : quelques éléments théoriques ;
- Manipulations:
 - TimeLineJS;
 - StoryMapJS.
- ! Lectures obligatoires
 - Robichaud et al. (2019)
- **?** Lectures supplémentaires
 - Perret (2022a)
 - Perret (2022b)
 - Drucker (2011)

[Séance 13:09/12] IA

- ! Important
 - Remise de l'examen final (écrit).
- ! Important

Cette séance est remplacée par les deux Ateliers *IA pour les SHS* :

- jeudi 11 septembre 2025 15:30 : Introduction : comment distinguer l'IA de ChatGPT ;
- jeudi 09 octobre 2025 15:30 : IA et la correction textuelle automatique : quels outils et quelles limites ?
- Lectures supplémentaires
 - Chagué (2022)
 - Smits & Wevers (2023)

[Séance 14 : 16/12]

! Important

Examen final.

Introduction au cours

Bienvenue!

Cette séance est dédiée d'une part à la présentation du cours, son déroulement et autres détails utiles, d'autre part à un (bref) panorama des Humanités Numériques.

Tour de table

Quelques questions:

- Quel est votre prénom ?
- Quel est votre parcours académique ?
- Qu'attendez-vous de ce cours ?
- Que savez-vous des Humanités Numériques ?

Déroulement du cours

- Cours en présentiel = présence requise ;
- Lectures obligatoires avant chaque séance :
 - Les lectures qui ne sont pas encore précisées le seront au moins une semaine avant la semaine concernée;
 - Les lectures obligatoires serviront de base à des discussions pendant le cours et sont à connaître pour les divers examens;
- Support/ressources disponibles à la fin de chaque cours (sur ce site web) :
 - Ce qui ne dispense aucunement la prise de note durant le cours ;
- Pendant les séances : partie théorique, manipulations, échanges.

Objectifs du cours

Cf. le Plan de cours.

HNU2000-A25 Introduction au cours

Présentation des séances

Cf. le Plan de cours.

Modalités d'évaluation

Cf. le Plan de cours.

[Séance 1 : 02/09] Introduction aux Humanités Numériques

Définitions

Comment définiriez-vous, en vos mots, les *Humanités Numériques*?

D'une part, les humanités numériques pourraient être définies comme l'application d'une méthode d'analyse informatique aux sciences humaines. En d'autres mots, l'approche des DH consiste à prendre en compte le fait que la puissance ne doit pas être limitée aux sciences dures, mais peut et doit aussi être employée pour des recherches en sciences humaines. D'autre part, les humanités numériques transcendent cet aspect technique et peuvent être pensées comme un regard global posé sur les changements culturels déterminés par le numérique ; en ce sens, les humanités numériques pourraient conduire à une sorte d'« humanisme numérique ». (Sinatra & Vitali-Rosati, 2014)

Domaine de recherche et d'enseignement au croisement de l'informatique et des lettres, des arts, des sciences humaines et des sciences sociales, visant à produire et à partager des savoirs, des méthodes et de nouveaux objets de connaissance à partir d'un corpus de données numériques.² (Commission d'enrichissement de la langue française, 2019)

• Quelques autres définitions ici ou ici.

Humanités...

- Traditionnellement : lettres classiques ;
- Aujourd'hui (et surtout en Amérique du nord) : les sciences humaines ;
- Littérature, philosophie, histoire, arts vivants, linguistiques, etc.

²Cette définition a été réutilisée sur le site de l'OQLF.

...Numérique(s)

- Représentation par nombre ;
- Discrétisation;
- Numérisation du monde ;
- Culture numérique.

Pour en savoir un peu plus sur l'encodage des caractères :

https://www.youtube.com/watch?v=MijmeoH9LT4&list=PLzH6n4zXuckqmf_xUcvU5caZVoctP2e hL

Les trois moments des Humanités Numériques

Lou Burnard (2012) distingue trois moments des humanités numériques :

env. 1970: Literary and Linguistic Computing

- Déjà vers la fin des années 1940 avec l'Index Thomisticus de R. Busa;
- Puissance de calcul des ordinateurs afin d'*automatiser* la création d'index, le repérage de concordances, le calcul de fréquences ... ;
- Ex.: British National Corpus, Thesaurus Linguae Graecae.

env. 1980: L'humanities computing

- Compréhension et maîtrise du programme informatique qui devient à proprement parler l'instrument d'une méthode de recherche et ce, dans une perspective interdisciplinaire;
- Ex. : la TEI (*Text Encoding Initiative*).

env. 1990 : Digital Humanities

- Apparition du Web → disponibilité des corpus numériques → nécessité de penser des interfaces pour consulter les corpus. Nouveaux rythmes de publication, modes de diffusions, formats.
- « On est passés de l'informatique comme outil au service de la recherche au numérique comme environnement global au sein duquel s'effectue la recherche. » (Debouy, 2025, p. 15)
- Ex.: la William Blake Archive.

Quelques repères historiques

- Aujourd'hui, l'informatique est partout → gain de temps, facilité de traitement; accès aux informations;
- Nouvelles questions : où sont stockées les données ? Quelle confidentialité ? Quelle est notre relation aux outils que nous utilisons tous les jours ?
- « Le citoyen du XXI^e siècle en est bien souvent réduit à être de plus en plus un consommateur et non un utilisateur averti et libre de ces technologies numériques. » (Debouy, 2025, p. 17)

Les prémices mécaniques

• II° s. av. J.-C. : le mécanisme d'Anticythère peut être considéré comme le premier calculateur analogique (connu). Il permettait de modéliser la course des astres grâce à une trentaine d'engrenages. Il faudra attendre près d'un millénaire pour les prochains systèmes comparables, dans les horloges de Moyen Âge. La reconstruction de ce mécanisme intéresse toujours les chercheurs aujourd'hui.



Figure 1: Le fragment principal de la machine d'Anticythère

• Env. 1640 : la *Pascaline* de Blaise Pascal, calculatrice mécanique capable d'additionner et de soustraire, conçue pour assister son père percepteur d'impôts.



Figure 2: Pascaline, conservée au musée des Arts et Métiers, Paris

• XVII^e s. – fin du XX^e s. : la règle à calcul, un instrument mécanique utilisé massivement jusqu'aux années 1970 pour faire des opérations arithmétiques et trigonométriques. Elle sera préférée à la Pascaline notamment en raison de son prix, sa facilité d'utilisation et de fabrication.

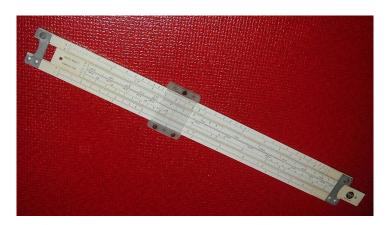


Figure 3: Règle à calcul scolaire, 10 pouces (Pickett N902-T simplex trig).

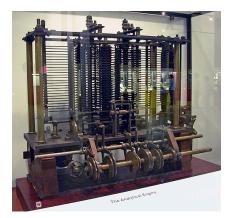
• 1801 : le métier à tisser de Jacquard, système mécanique programmable par l'utilisation de cartes perforées, introduit l'idée d'un programme externe pour automatiser des opérations.

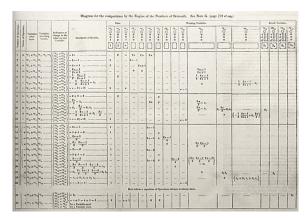


Figure 4: Métier à tisser de J.-M. Jacquard

Les concepts fondateurs

• Env. 1830 : Charles Babbage conçoit la machine analytique, un ordinateur programmable universel avant l'heure, inspiré du métier à tisser de Jacquard. Elle ne sera jamais construite faute de moyens techniques et financiers, mais ses plans étaient fonctionnels. Ada Lovelace y voit la possibilité d'écrire de véritables programmes (une suite d'instructions à mémoriser, exécuter et produire un résultat) et imagine déjà que la machine pourrait manipuler n'importe quel type de symbole — anticipant l'usage des ordinateurs pour le texte, la musique ou les arts: elle conçoit le premier algorithme en plus d'élargir la portée conceptuelle de la machine.





- (a) Prototype (1871) non terminé de la machine analytique(a) « Programme » de calcul des nombres de Bernoulli dans de Babbage une note d'A. Lovelace (1843)
 - 1890 : Développement (par Herman Hollerith) des machines à cartes perforées pour le recensement américain, à l'origine d'IBM.

• 1936 : Alan Turing formalise, dans son article *On Computable Numbers*, la notion de *machine universelle*, un modèle théorique capable d'exécuter tout algorithme. Turing est ainsi une figure primordiale de l'histoire de l'informatique. Durant la Seconde Guerre mondiale, il contribue également au déchiffrement des codes secrets allemands (notamment *Enigma*)³.

Vers l'ordinateur moderne

• 1945 : John von Neumann formalise l'architecture du *programme enregistré*, où instructions et données sont stockées dans une même mémoire.

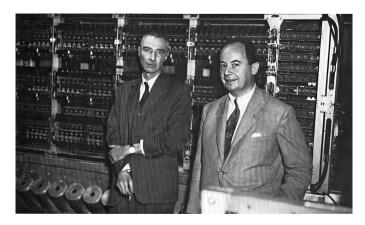


Figure 7: Oppenheimer et von Neumann à l'inauguration en 1952 de l'ordinateur construit pour l'*Institute for Advanced Study*.

 Années 1950 : IBM commercialise ses premiers ordinateurs ; FORTRAN⁴ (FORmula TRANslator) devient le premier langage de haut niveau utilisé massivement pour le calcul scientifique et l'ingénierie.

L'informatique et les sciences humaines

• 1949 : le jésuite Roberto Busa, avec IBM, lance le projet de l'*Index Thomisticus* (analyse assistée par ordinateur de l'œuvre de Thomas d'Aquin), considéré comme l'acte fondateur des humanités numériques.

Démocratisation et réseau

 Années 1960–1970: naissance d'Internet, à la suite de projets militaires et universitaires (comme AR-PANET, financé par le Department of Defense américain) qui mettent au point un réseau de communication décentralisé, capable de transmettre des paquets d'informations entre ordinateurs.

³C'est sur cet épisode que porte le film *The Imitation Game* (2014).

⁴Un bac à sable est disponible ici.

• Années 1980 : diffusion des micro-ordinateurs personnels, qui ouvrent l'informatique à la recherche, l'enseignement et le grand public.



Figure 8: IBM PC 5150 avec clavier

• 1990 : Tim Berners-Lee (avec la collaboration de Robert Cailliau) invente le World Wide Web au CERN, rendant possible la circulation mondiale des textes numériques.

Note

Les concepts d'Internet et du Web sont complètement différents! Le Web repose sur Internet, mais il en est une application.



(a) Tim Berners-Lee



(a) Robert Cailliau

• Années 1990-2000 : grands projets de bibliothèques numériques (Gutenberg, Perseus, Gallica) et standardisation de l'encodage de texte (TEI).

• 1998 : Google est créé par Larry Page et Sergey Brin. Il s'agit d'un moteur de recherche qui classe les pages Web selon leur popularité (PageRank).

Vers les humanités numériques contemporaines

- Années 2000-2010 : massification du numérique en SHS : OCR → HTR, bases textuelles, textométrie, fouille de texte ;
- Années 2010-2020 : montée en puissance de l'IA et des grands modèles de langage, offrant de nouvelles méthodes d'analyse, mais aussi de nouveaux enjeux critiques.

[Séance 2 : 09/09] COMPRENDRE son environnement

Retour sur la lecture

- Lectures obligatoires
 - Sinatra & Vitali-Rosati (2014)

Terminal, ligne de commande

La ligne de commande

- *Interfaces graphiques* = couche superposée à une interface de communication fondamentale : la commande ;
- Le *Terminal* que nous allons manipuler fait référence au logiciel avec lequel nous pouvons taper des lignes de commandes afin d'interagir avec le *système d'exploitation*;
- Shell : interpréteur de commandes qui sert d'intermédiaire entre les utilisateur·ices et le système d'exploitation ;
 - Différents shells existent, le shell par défaut de la plupart des distributions Linux & MacOS X se nomme Bash.
- Les interfaces graphiques "traduisent" en lignes de commande les opérations que nous faisons (mais nous le cachent).
- Avantage : la ligne de commande est + rapide et + sûre.

OS: « logiciel faisant office d'interface [(graphique)] entre les composants informatiques et l'utilisateur, qui assure notamment l'exécution des programmes, l'allocation des ressources matérielles et le contrôle des opérations d'entrée-sortie sur les périphériques. » (OQLF 2022).

Exemple

Création d'un répertoire (= Dossier) sur le Bureau puis y déplacer un fichier que l'on vient de créer :

- 1. Clic droit, Dossier > Nouveau → une fenêtre apparaît;
- 2. Saisir le nom du répertoire (Test) > cliquer sur OK;
- 3. Déplacer le fichier exemple.pdf en le faisant glisser avec le bouton gauche de la souris sur l'icône du répertoire puis relâcher le bouton.

Création d'un répertoire (= Dossier) puis y déplacer un fichier que l'on vient de créer avec la ligne de commande :

```
1. ouvrir le terminal;
```

- 2. mkdir Test \rightarrow *Enter*;
- 3. mv exemple.pdf Test \rightarrow *Enter*.

La ligne de commande, plus rapide et plus sûre ?

- Interface graphique = surcouche logicielle → ralentit le système d'exploitation;
- Interface graphique = logiciel complexe → *bugs*;
- Interface graphique ne contient pas *tous* les équivalents qu'a la ligne de commande certaines opérations ont été laissées de côté et ne pourront être effectuées par l'utilisateur-ice;
- Séquence de lignes de commande, e.g.: mkdir Test; mv exemple.pdf Test;
- Présence de *wildcards* dans les lignes de commande, permettant de déclencher des opérations complexes sur un grand nombre de fichiers, *e.g.*: mv *.pdf Test déplacera tous les fichiers PDF dans le répertoire Test.

Testons

- Linux : chercher *terminal* dans les *outils système* (raccourci : CTRL+ALT+t);
- MacOS: chercher terminal dans les applications (toutes les commandes ne sont pas présentes);
- Windows: installer Cygwin.

Où sont les fichiers sur un ordinateur ?

- Pourquoi c'est important ?
 - Interactions facilités avec les différents fichiers constituant un projet;
 - Liens (relatifs et/ou absolus) entre fichiers.

Arborescence

- Chaque système d'exploitation a son propre système d'organisation ;
- Il faut comprendre les systèmes de chemin ;
- Tout est arborescence

Figure 11: Arborescence sur un système Ubuntu

- Votre disque dur = sorte de grande boîte où se trouvent d'autres boîtes (= répertoires/dossiers);
- *Racine* = boite principale contenant toutes les autres ;
 - Mac, Linux:/;
 - Widows: $C: \$.

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

Comprendre le fonctionnement global de l'arborescence permet de se repérer dans son ordinateur.

Nommer ses fichiers

- Quelques bonnes pratiques :
 - Les noms de fichier ne doivent pas contenir des caractères spéciaux, ni des accents, ni des espaces :
 l'ordintateur ne les reconnaît pas immédiatement, il doit les convertir, ce qui peut provoquer des erreurs ;
 - Les noms des fichiers doivent être significatifs ;
 - Les noms des fichiers ne doivent pas dépasser 80 caractères ;

- Faites attention aux extensions de fichiers! Il est préférable de ne pas les masquer: information primordiale pour la machine ET les utilisateurs;
- Exemple: 2025-09-09-hnu2000-seance02.md.

Chemins relatifs et chemins absolus

Il est possible d'indiquer à la fois un chemin relatif, donc sans commencer par une barre oblique, ou en commençant par une barre oblique pour un chemin absolu.

Chemin absolu	Chemin relatif
/Users/Mathilde/Documents/Projet/index.html	./Documents/Projet/index.html
Emplacement complet à partir de la racine du	Chemin vers un fichier à partir du répertoire actul
système de fichiers	(pwd)
Toujours valide, qu'importe le répertoire	Change selon où je me situe dans mes fichiers
Commence nécessairement par /	Ne commence pas par /

Astuces:

- / : on descend d'un niveau dans la hiérarchie des répertoires ;
- . . : on remonte vers un répertoire parent.

Les premières commandes

Quand je lance mon terminal, une invite de commande (*prompt*) apparaît,

```
e.g. mathilde@crcen-mathilde-xps-13-7390:~:
```

- l'utilisateur mathilde est connectée sur l'ordinateur crcen-mathilde-xps-13-7390;
- le ~ indique qu'on se trouve dans le *home directory*.

Où suis-je?

• pour savoir où l'on se trouve : pwd



pwd nous retourne un chemin absolu: le chemin indiqué commence en effet par une barre oblique.

Déplacer un fichier

- La commandemv <source> <destination> permet de renommer ou de déplacer un fichier (dans le terminal, c'est la même chose!);
- mv mon-fichier.txt fichier.txtrenommelefichiermon-fichier.txtenfichier.txt;
- mv /home/utilisateur/hnu2000/mon-fichier.txt /home/utilisateur/Documents/ déplace le fichier mon-fichier.txtdans le dossier Documents (il doit exister préalablement!).

Naviguer dans ses fichiers

- pour naviguer : cd suivi du chemin,
- exemple:
 - cd Documents/cours/hnu2000;
 - cd : retour au *home directory*.

Note

Il est possible d'indiquer à la fois un chemin relatif, donc sans commencer par une barre oblique, ou en commençant par une barre oblique pour un chemin absolu.

Remarquez le changement de l'invite après l'exécution de la commande.

Lister des fichiers

- pour lister les fichiers : ls ;
 - ex.:ls Test/>fichier.txt;
- pour lister les fichiers, dont les fichiers cachés : 1s -a.

Créer un dossier

• mkdir + le nom du dossier (*MaKe DIRectory*).

Créer un fichier

• touch + le nom du fichier.

Supprimer un fichier

- rm mon-fichier.md;
- rm -R mon-dossier supprime le dossier et tout ce qu'il contient.

Afficher le contenu d'un fichier

• cat mon-fichier.md.

Comprendre une commande

• pour comprendre à quoi sert une commande : man nom-de-la-commande (ex. : man rm).

Quelques réflexes

- La touche TAB complète une commande ;
- La flèche du haut parcourt les dernières commandes utilisées ;
- CTRL + R pour rechercher une commande précédemment tapée ; la recherche se mettra à jour au fur et à mesure que vous tapez les premières lettres de la commande ;
- CTRL + L ou clear pour repartir sur un nouveau terminal.

Exercice

- créez un dossier HNU2000 (dans un dossier Cours par exemple) contenant un sous-dossier 02-exercice contenant lui-même un fichier exemple.txt;
- naviguez dans le dossier 02-exercice;
- revenez dans le dossier Cours;
- déplacez le fichier exemple.txt dans le dossier HNU2000;
- renommez le fichier exemple.txt en 02_exercice1.txt;
- supprimez le dossier 02-exercice.

Quelques commandes avancées

- history: Pour afficher et parcourir l'historique des commandes que vous avez utilisées (seulement les commandes, pas les résultats de ces commandes);
- less + fichier : Affiche le contenu d'un fichier page à page (similaire à cat, plus lisible),
 - pour quitter : q!
- cp <source> <destination> copie un fichier ou un dossier :

- cp HNU2000/texte.txt Documents/divers/textes/texte.txt; - cp -r textes Documents/divers;
- we permet de compter le nombre de lignes, de mots et de caractères dans un fichier texte;
- find permet de chercher dans les noms de fichiers et dossiers, et dans les fichiers eux-mêmes;
- && pour enchaîner les commandes.



Testez la commande suivante : curl http://wttr.in/Montréal\?lang\=fr. Nous y reviendrons plus tard dans la session.

Bash

- Interpréteur en ligne de commande d'Unix (= Linux, Mac);
- Un script Bash (ou shell) est un fichier (en .sh) comportant plusieurs instructions/commandes;
- Celles-ce seront exécutées les unes après les autres ;
- → Automatisation des tâches.

Exercices

Renommer tous ses fichiers automatiquement

Si vous désirez ajouter HNU ou vacances 2025 devant chaque nom de fichier dans un dossier donné, créez un script :

- Créez (en ligne de commande), un fichier intitulé rename.sh;
- Ajoutez-y le script suivant :

```
#!/bin/bash
# Ajoute "hnu_" devant chaque fichier

for file in *; do
    if [ -f "$file" ]; then
        mv "$file" "hnu_$file"
    fi
done
```

- Créez quelques fichiers de test (touch test1.txt test2.txt);
- Faites tourner le script : bash rename.sh.

Le script:

Changer l'extension automatiquement

```
#!/bin/bash
# Change l'extension .md en .txt pour tous les fichiers

for file in *.md; do
    if [ -f "$file" ]; then
        mv "$file" "${file%.md}.txt"
    fi
done
```

Générer du texte

Testez plusieurs scripts Bash. Pour ce faire, créez des fichiers nom-du-fichier.sh, ajoutez-y le contenu indiqué ci-dessous, et faites les tourner en tapant, dans votre terminal, bash nom-du-fichier.sh:

```
echo "texte";
valeur="texte";
echo ${valeur}

valeur="texte";
echo "je suis du ${valeur}"
```

Jouer avec les commandes de base

• Afficher la date :

```
horaire=`date`
echo "Jounal de bord du $horaire"
```

• Afficher la situation :

```
lieu=`pwd`
echo "Je me trouve toujours à bord du $lieu"
```

• Demander mon identité : (hors interface en ligne)

```
identite=`whoami`
echo "Je me prénomme $identite"
   • Aligner du texte :
T[0]='texte1'
T[1]='texte2'
echo ${T[0]} ${T[1]}
   • Création d'une interaction :
echo "To be or not to be?"
             select i in être non-être; do
                 if [ "$i" = "être" ]; then
                     echo "Sois";
                     break
                 elif [ "$i" = "non-être" ]; then
                     echo "Tant pis"
                     break
                 else
                     echo "Pas de troisième choix dans ce script"
                 fi
        done
```

Astuce

D'autres exemples ici.

WYSIWYM >< WYSIWYG?

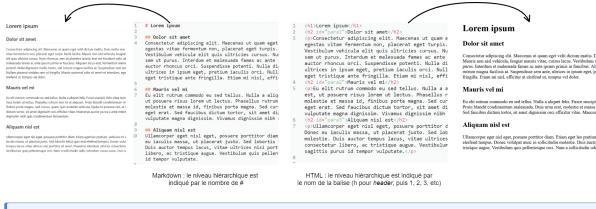
Deux grandes familles de systèmes de saisie de documents textuels :

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

Le logiciel de traitement de texte Word est-il WYSIWYM ou WYSIWYG?

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).





Si nous avons le temps, lecture en classe de ce billet.

Mettre en forme =/= structurer

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

VSCodium: installation et manipulations

C'est quoi Codium?

• Version open source de Visual Studio Code (VS Code)

VSCodium is a community-driven, freely-licensed binary distribution of Microsoft's editor VS Code.

• VS Code : éditeur de code développé par Microsoft. Gratuit, léger et puissant. Largement utilisé par la communauté des développeur-ses pour faire de la programmation.

Extensions

Suivre ce tutoriel (pour installer French Language Pack for VS Code, French - Code Spell Checker et LTeX, à configurer en français).

[Séance 3 : 16/09] CHERCHER: corpus et ressources numériques

Retour sur les lectures et écoutes

- Lectures obligatoires
 - France Inter (s. d.)
 - Huma-Num (2021)

Une histoire d'internet et du Web

Nous avons évoqué, lors de la première séance, Internet et le Web. Qu'en avez-vous retenu ? S'agit-il de la même chose ?

Ce ne sont pas tant les « nouvelles technologies » en général, mais le réseau en lui-même qui a bouleversé notre rapport à la connaissance. (Mille, 2014)

Origines

- 1962-1968 : ARPAnet
- 1969-1978 : Internet
- années 1970 et 1980 : accès aux machines
- 1984 : un réseau fonctionnel

Plus de détails dans (Mille, 2014).

Principes: TCP/IP

TCP/IP:

- TCP (Transmission Control Protocol);
- IP (*Internet Protocol*).

Une pile de protocoles :

- physique : les câbles (cuivre ou fibre), éventuellement radio ;
- liaison : Ethernet ou Wireless Ethernet (Wifi), il s'agit de déterminer comment les paquets sont acheminés ;
- réseau : c'est la partie IP qui permet d'acheminer des paquets en donnant des adresses à toutes les machines connectées sur un réseau ;
- **transport** : c'est la partie TCP, pour transférer les informations découpées en paquets et reconstituées en vérifiant qu'il ne manque rien ;
- application : c'est là qu'apparaissent les services qu'on connaît : le Web avec HTTP, les emails avec SMTP ou IMAP, ou encore le FTP pour partager des fichiers.



Pour en savoir plus sur Internet et ses protocoles, voici une série de vidéos très complètes : https://iletai tunefoisinternet.fr/.

Expérimentation

Découvrir la route pour accéder à un serveur :

- utiliser tracepath (ou traceroute) en ligne de commande
- lancer la commande tracepath umontreal.ca ou traceroute umontreal.ca
- analyser les résultats

Le Web

- Internet ≠ Web
- le Web = une application d'Internet
- un protocole (HTTP) et des langages (HTML/CSS/JavaScript)

Le Web est une application d'Internet permettant de publier et de consulter facilement des informations.



Le CERN a restauré la première page du Web.

Exercice

- comment lire une page web via un autre outil qu'un navigateur ?
- utiliser la commande curl dans le terminal ;
- exemple 1:curl https://mathildevrst.github.io/HNU2000-A25/Plan-Cours.html;
- exemple 2: curl https://theread.me/raw-permalinks-for-accessibility/.

Les navigateurs et moteurs de recherche

Au principe d'autorité qui a fait la force du PageRank, Google substitue de plus en plus un principe d'efficacité qui renvoie de manière toujours plus appropriée vers l'internaute les choix que l'algorithme a appris de ses comportements. (Cardon, 2013)

- Navigateur ≠ moteur de recherche⁵!
 - Navigateur : application, installée sur l'ordinateur, qui permet d'afficher les informations contenues sur les sites Web;
 - Moteur de recherche : application qui permet de localiser des informations sur le Web, [...] outil
 devenu indispensable à partir du moment où les ressources Web, devenues trop nombreuses, ont
 rendu difficile l'indexation des contenus en ligne.

Quels navigateurs utilisez-vous? Pourquoi?

Le Web, au commencent étaient les annuaires

- Dans les années 1990, les premiers répertoires (Yahoo!, DMOZ) proposaient une liste classée de sites web sous forme d'une arborescence thématique;
- La démarche est différente d'aujourd'hui : on naviguait dans une hiérarchie, on ne formulait pas encore une requête.
- Avantages : navigation linéaire par thématiques, parcours dans une arborescence logique, aperçu potentiellement exhaustif de l'existant.
- Inconvénients : parfois une seule entrée pour un résultat qui concerne plusieurs thématiques, recherche fastidieuse ; explosion du Web rendant le modèle vite obsolète.

⁵Les deux définitions ci-dessous sont issues de (Debouy, 2025, p. 125).



Figure 12: Yahoo! en 2001

Explorer, indexer, classer

Un moteur de recherche recense et référence des pages Internet : pour cela, il utilise un logiciel, un robot d'indexation (dits *crawlers* ou *spiders*) qui parcourt l'ensemble du Web :

- 1. Exploration : exploration régulière du Web par le *crawler* qui suit tous les liens trouvés puis analyse les pages jugées intéressantes ;
- 2. Indexation : extraction de mots-clés des pages visitées et conservation d'une copie (= *cache*) de ces pages ;
- 3. Classement : classement, par le moteur de recherche, des résultats en fonction de leur pertinence. C'est là qu'interviennent les algorithmes, qui tiennent compte par exemple de la popularité d'une page (combien d'autres sites font un lien vers elle), mais aussi de centaines d'autres critères que Google garde secrets.

Sundar Pichai, président-directeur général de Google, tente d'explique au Congrès Américain comment fonctionne Google

https://www.youtube.com/watch?v=8fP95zyPePU

Quelques moteurs de recherche

Google

- Le plus utilisé, merci le *PageRank*, algorithme mathématique utilisé pour déterminer la pertinence des pages Web;
- Attention aux premiers liens qui peuvent être sponsorisés ;

- Personnalisation des résultats : deux utilisateurs effectuant la même requête n'obtiendront pas les mêmes résultats (prise en compte des recherches et activités passées, ainsi que du profil des utilisateurs (situation géographique, âge, langue, etc) (Ochigame & Ye, 2021));
- Quid des données personnelles ?

Alternatives

Quelques moteurs de recherche n'enregistrent pas les recherches effectuées et n'exploitent pas les données personnelles à des fins publicitaires ; ne filtrent pas les résultats de recherche en fonction des profils des utilisateurs → neutralité des recherches :

Qwant	DuckDuckGo	Startpage
"Le moteur de recherche dont vous êtes l'utilisateur, pas le produit"	"Le moteur de recherche qui ne vous espionne pas"	"Le moteur de recherche le plus privé au monde"
Exploite simplement les banques de données réalisées par les	Métamoteur qui interroge les banques de données des autres	← Idem
<i>crawlers</i> Français & Vie privée	moteurs de recherche Gain de temps & Protection des données	← Idem

Google Scholar

- Moteur de recherche généraliste ≠ Moteur de recherche d'articles scientifiques ;
- Généralement, accès au texte intégral;
- Mais, pas de possibilité de tri/filtre des résultats ;
- Pas de vérification du caractère scientifique des articles ;

Isidore

- Projet de recherche, 2011 < TGIR Huma-Num;
- objectifs : mettre à disposition un outil de recherche adapté aux sciences humaines ;
- ambition : proposer des fonctionnalités pour faciliter la recherche documentaire ;
- S'appuie sur le principe du Web de données ;
- Se distingue des autres moteurs de recherche :
 - Moissonage ciblé et des métadonnées et données scientifiques structurées selon des standards internationaux et disponibles en accès libre
 - Indexation des données non structurées et structurées ;
 - Normalisation des métadonnées et enrichissement des données en s'appuyant sur des référentiels reconnus;

- Mise en évidence des sources de données indexées ;
- Récupération des actualités ;
- Algorithmes basés sur des référentiels selon des techniques documentaires ≠ PageRank.

Exercice

- constituez une bibliographie dans Isidore.science sur les humanités numériques ;
- pour cela vous devez vous créer un compte Isidore (et donc HumanID);
- utilisez les fonctionnalités de la recherche avancée ;
- utilisez les fonctions de tris et de facettes ;
- partagez avec moi votre « bibliothèque ».

Discussion : Que manque-t-il à Isidore ? Quels sont ses défauts ?

Google et le PageRank

En 1998, Google introduit le PageRank, qui attribue une importance à une page selon le nombre et la qualité des liens qui pointent vers elle (= système de mesure quantitative de popularité d'une page web). C'est inspiré par la mesure des articles académiques (Science Citation Index).

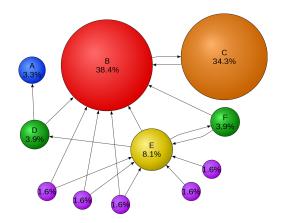


Figure 13: Schéma du PageRank

Cela fonde un principe d'autorité : les pages citées par beaucoup d'autres deviennent les plus visibles. Progressivement, Google combine ce principe avec d'autres critères (localisation, personnalisation).

Les évolutions des moteurs de recherche

Il y a plusieurs évolutions majeures des moteurs de recherche ces dernières années :

- Interface utilisateur (UI) : simplification radicale de la page Google, disparition progressive des options avancées ;
- Aide à la formulation : autocomplétion, suggestions liées aux recherches fréquentes ;
- Web sémantique et Knowledge Graph (2012–) : affichage de données structurées (fiche Wikipédia, horaires, météo,...) directement dans les résultats ;
- Personnalisation : traçage des comportements, adaptation aux historiques et aux contextes ;
- SEO (Search Engine Optimization): pratiques d'optimisation par les sites pour "plaire" à Google;
- IA générative (2023–): moteurs qui ne se contentent plus de lister des pages, mais produisent directement des résumés de réponse (Google SGE, Perplexity).

Phénomène de simplification :

- Suppression des options → interface minimale, recherche "intuitive";
- Autocomplétion → orientation subtile des requêtes ;
- Pertinence des premiers résultats → effet d'écrasement : très peu d'utilisateurs consultent la deuxième page ;
- Affichage de données liées (web sémantique) → l'utilisateur n'a plus besoin de cliquer sur les sites sources;
- IA générative → bascule majeure : le moteur devient assistant ou agent conversationnel, qui filtre et reformule le Web au lieu de seulement l'indexer;
- Tendance globale → l'information va de plus en plus vers l'utilisateur, au risque de réduire sa diversité d'accès.



Figure 14: Outil de réservation SNCF, capture d'écran, 18 août 2025

Vade-Mecum d'une recherche efficace

- 1. Commencer simple
- Tapez quelques mots-clés principaux de votre sujet ;
- Utilisez des mots précis plutôt que des phrases longues ;
- Google ignore généralement les mots très fréquents (le, la, de...).
- 2. Utiliser les guillemets pour les expressions exactes
- "expression exacte" → Google recherche ces mots dans cet ordre précis;
- e.g.: "humanités numériques" ne retournera que les pages contenant exactement cette expression.

- 3. Limiter la recherche à un site ou un domaine
- site:lesite.com requête → Recherche uniquement sur ce site;
- e.g: site:umontreal.ca "humanités numériques" → pages sur l'Université de Montréal concernant les humanités numériques.
- 4. Chercher un type de fichier spécifique
- filetype:pdf requête → Cherche des fichiers PDF uniquement;
- e.g.:filetype:pdf "édition numérique" → documents PDF sur l'édition numérique.
- 5. Chercher dans le titre d'une page
- intitle:mot → Page dont le titre contient ce mot;
- e.g.:intitle: "humanités numériques" → pages dont le titre contient l'expression exacte.
- 6. Les opérateurs
- AND⁶, OR, NOT permettent de préciser la recherche. Des parenthèses servent à grouper des termes ou des opérateurs;
 - e.g.: "humanités numériques" AND "édition" -cours \rightarrow pages sur l'édition dans les humanités numériques mais sans cours ;
 - e.g. : ("humanités numériques" OR "digital humanities") AND "édition" → pages qui contiennent soit "humanités numériques", soit "digital humanities", mais qui contiennent aussi "édition".
- ?: Remplace un caractère;
 - e.g.: econom? recherche économie, economia.
- * : Élargit la recherche aux variantes d'un même mot ;
 - e.g.: pers* recherche personnage, personnalité, persona,...
- > < = :: précise une période chronologique.



Astuce

Les formulaire de recherche avancée des bibliothèques utilisent ces opérateurs!

⁶Si les mots-clés ne sont pas reliés par un opérateur, la plupart des outils de recherche considèrent que les termes sont reliés par AND.

Exercice

Dans votre moteur de recherche favori:

- Cherchez une expression exacte;
- Cherchez là uniquement sur un site spécifique ;
- Cherchez un type de fichier particulier;
- Cherchez des pages dont le titre contient un mot-clé;
- Combinez plusieurs opérateurs pour affiner la recherche ;
- Cherchez toutes les pages qui parlent d'humanités numériques et d'édition sur le site web de l'Université de Montréal.

Une alternative aux moteurs de recherche : les flux RSS

- Really Simple Syndication;
- Ressources dont le contenu est mis à jour de manière automatique selon les modifications d'un site Web;
- On s'abonne à un flux en l'ajoutant dans un agrégateur qui vérifiera automatiquement si de nouveaux contenus ont été mis en ligne ;
 - Parmi les agrégateurs, des applications Web (Netvibes, Feedly, Inoreader, FreshRSS,...) ou des extensions de navigateur (Feedbro,...);
- Avantage : on suit les actualités d'un site sans devoir s'y redre manuellement.

[Séance 4 : 23/09] ORGANISER & BIBLIOGRAPHIER: métadonnées et formats

Retour sur les lectures

- Lectures obligatoires
 - Collectif (2018)

Les métadonnées

- Données sur des données ;
- Sont des métadonnées les notices bibliographiques, le format d'un fichier, la date, la géolocalisation d'une photo,...

Composé à l'aide du préfixe grec *meta*, le terme désigne à proprement parler les données à propos des données : elles servent à référencer, identifier et partager un document. Elles indiquent notamment comment, quand, où et par qui les données ont été recueillies, mentionnent leur disponibilité et leur mode de distribution, leur taille, etc. (Debouy, 2025, p. 160)

Pourquoi les métadonnées ?

- Trouver : elles rendent les données repérables (moteurs de recherche, catalogues) ;
- Comprendre: elles donnent du contexte (qui a produit la donnée, quand, dans quelles conditions);
- Réutiliser : elles précisent les droits, les formats, les relations avec d'autres données ;
- Préserver : elles facilitent l'archivage et l'interopérabilité à long terme.

Les différents formats textuels

Les formats : un peu d'étymologie

• 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**.

Vers une définition du format

- Ce qui donne une forme et un ensemble de caractéristiques à un ensemble d'informations (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 ;
- interopérabilité.

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).

En informatique, au niveau le plus fondamental, tout est exprimé dans un alphabet numérique binaire : 0 et 1. Un format, c'est une convention qui établit la correspondance entre une certaine succession de 0 et de 1, et quelque chose d'autre : par exemple une couleur, ou une lettre de l'alphabet, ou la position d'un pixel sur un écran. L'expression « format texte » désigne une catégorie de formats pour lesquels le contenu en binaire des fichiers encode des caractères textuels uniquement. (Arthur Perret)

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 (métadonnées → identification);
 - structuration (normalement) rigoureuse (balisage sémantique, non ambigu);
 - pérennité;
- Les formats utilisés doivent répondre à ces exigences, permettant notamment d'expliciter la fonction de chaque élément du document (titre, citation, métadonnées, références bibliographiques,...)

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.

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

Le format n'est pas neutre!

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

Zoom sur le libre

- Discussion : Qu'en savez-vous ? Intérêts, avantages, inconvénients,... ?
- Quid des productions scientifiques des chercheurs publiées dans des revues scientifiques en limitant l'accès ?
- Question du partage et de la diffusion des connaissances scientifiques ;
- Question de la possibilité du lien entre les données

En agrégeant des contenus, en lian les publications avec les données de la recherche et à bien d'autres objets, les techologies du Web sémantique permettent la construction de nouveaux savoirs. (Debouy, 2025, p. 153)

Le cas des revues

- Depuis les années 2000, développement des revues en accès ouvert : plus de 10.000 à ce jour!
 - ex. : OpenEdition : portail (français) de ressources électroniques en SHS ;
 - ou HAL: archive ouverte pluridisciplinaire.

Différentes voies possibles :

1. green open access:

- autoarchivage permettant l'accès libre et gratuit aux articles scientifiques ;
- si l'article a déjà été publié par un éditeur :
 - celui-ci le libère
 - mais, souvent, période d'embargo afin d'inciter les utilistaeurs à continuer à souscrire des abonement

2. gold open access:

- l'auteur (ou son institution d'appartenance) paie pour sa publication ;
 - // auteur-payeur

3. les épi-revues:

- Voie dorée par l'hébergement de revues en open access, et
- Voie verte via le processus de soumission des articles qui se fait sur une archive ouverte ;
- Pas de frais d'accès pour les lecteurs ni de frais de publication pour les auteurs ;
- Les articles sont immédiatement consultables en accès ouverts (avant même l'évaluation par les pairs);
- Finalement, validation de l'article pour publication dans la revue : nouvelle version soumise sur l'archive + label de la revue ;
- Ex.: https://open-research-europe.ec.europa.eu/articles/5-16/v1.

Qu'est-ce qu'une bibliographie ?

L'importance de structurer ses références bibliographiques

- Communauté scientifique & standards : inscrire votre travail dans les standards scientifiques d'une communauté ;
- Éthique : qui réfère bien respecte le travail cité ;
- Pérennité : pour retrouver vos sources médiation d'un travail de sources à la base de votre recherche ;
- Esthétique : une bibliographie imparfaite fait tâche dans un travail de recherche ;
- Personnel : vous avez le choix de structurer ces références selon le type de style bibliographique.

Style de citation vs style bibliographique

Style de citation	Style bibliographique
règles utilisées pour citer des sources dans le texte	présentation complète des sources dans une
	bibliographie ou une liste de références
APA: (Dupont, 2020, p. 45)	APA : Dupont, J. (2020). Titre du livre. Éditeur.
Vancouver:(1)	Vancouver : Dupont J. Titre du livre. Lieu de
	publication: Éditeur; 2020. 1000p.

Répertoire des styles : https://www.zotero.org/styles

Exemple

APA 7th - anglophone	Merzeau, L. (2013). Éditorialisation collaborative d'un événement.
	Communication et organisation. Revue scientifique francophone en
	Communication organisationnelle, 43, Article 43.
APA 7th - francophone	Merzeau, L. (2013). Éditorialisation collaborative d'un événement.
	Communication et organisation. Revue scientifique francophone en
	Communication organisationnelle, (43), 105-22.
Modern Language	Merzeau, Louise. "Éditorialisation collaborative d'un événement."
Association 9th edition	Communication et organisation. Revue scientifique francophone en
	Communication organisationnelle, no 43, June 2013, pp. 105-22.
Chicago Manual of Style	Merzeau, Louise. 2013. "Éditorialisation collaborative d'un événement."
17th edition	Communication et organisation. Revue scientifique francophone en
	Communication organisationnelle, no 43 (June): 105-22.

Ensemble de données!

Comment créer une bibliographie?

- Plusieurs milliers de styles bibliographiques : https://www.zotero.org/styles (choisir la *norme* que l'on va utiliser) ;
- Exemple : « Université de Montréal APA (Français Canada) » ;
 - Pour modifier le style par défaut de Zotero : Édition > Préférences > Citer ;
- Création d'une bibliographie : chaque référence ou groupe de références peut être récupéré sous différents formats pour être intégré dans un document via un traitement de texte par exemple.

Plus d'informations: https://www.zotero.org/support/fr/creating_bibliographies

Zotero

Qu'est-ce que Zotero?

- Un logiciel et un service qui permet de gérer des références bibliographiques : conserver, organiser, citer et créer des bibliographies ;
- Un logiciel libre, et gratuit;

- Un outil qui peut être collaboratif;
- Un logiciel + une extension de navigateur pour récupérer des références.

Fonctionnalités

- Collecter : collecter des références bibliographiques en ligne et les conserver ;
- Organiser : organiser ses références en collections et sous-collections, et en attribuant des tags et des marqueurs ;
- Citer : citer ses références facilement et les exporter/intégrer dynamiquement dans des documents ;
- Synchroniser : synchroniser ses références avec un compte en ligne ;
- Collaborer : échanger ses références avec d'autres utilisateurs.

Un logiciel libre

Zotero est logiciel libre qui peut être téléchargé puis installé sur tous les ordinateurs.

Il faut également utiliser un connecteur pour récupérer des références sur le web puis faire le lien avec le logiciel qui les conserve.

Utilisation générale

- Récupérer des références depuis un navigateur web (grâce à l'extension Zotero);
- Une icône apparaît pour enregistrer une référence (plusieurs types de documents);
- La référence est enregistrée dans le logiciel Zotero sous la forme d'une notice ;
- Il est possible de compléter la notice s'il manque des informations ;
- La synchronisation permet de sauvegarder les données en ligne (pour cela il faut créer un compte Zotero).

Fonctionnalités proposées avec un compte

- Synchronisation : vos références sont sauvegardées et disponibles en se connectant depuis n'importe quel ordinateur ;
- Sauvegarde;
- Interface web: un peu moins pratique que le logiciel;
- Afficher ses références ;
- Créer ou participer à des groupes ;
- Etc.

À quoi sert un groupe?

Un groupe est une collection à laquelle plusieurs personnes peuvent participer :

- Partager des références ;
- Ouvrir les contributions ;
- Afficher les contributions ;
- Plusieurs rôles dans un même groupe : consulter, créer, modifier, administrer ;
- Le groupe peut être public avec des permissions plus ou moins ouvertes

Installation

- Installer le logiciel Zotero en suivant les recommandations du site web officiel;
- Installer l'extension pour Firefox/Chrome/Safari;
- Créer un compte en ligne;
- Connecter votre logiciel à ce compte (Édition > Préférence > Synchronisation).

De nombreuses bases de données compatibles

De nombreux catalogues, bases de données, revues en ligne, journaux, blog, réseaux sociaux proposent des références :

- Sudoc;
- Gallica;
- OpenEdition, revues.org, Hypothèses;
- Érudit;
- Google Livres, Google Scholar, YouTube;
- Internet Archive;
- Cairn.info;
- etc.

Description d'une utilisation classique

Prérequis : logiciel installé et extension installée (par exemple dans Firefox).

- Ouvrir le logiciel (Zotero Standalone);
- Rechercher un document sur le web;
- Enregistrer la référence via l'extension du navigateur ;
- Vérifier dans le logiciel si la référence convient (et si elle est placée au bon endroit);
- Compléter les informations.

Quelques points importants

- Pas besoin d'être en ligne pour utiliser Zotero (mais vous ne pourrez pas afficher de pages web ou synchroniser vos collections);
- Les fonctions de Zotero sont nombreuses, n'hésitez pas à les tester et à les utiliser!
- Zotero permet de conserver et de gérer des références de pages web, mais ce n'est pas sa fonction première;
- N'oubliez pas d'ouvrir le logiciel Zotero pour enregistrer vos références.

Collecter

- Via la navigateur en cliquant sur l'icône : Zotero récupère les métadonnées de la page web (page web qui présente un document);
- Si le texte intégral est disponible en PDF, Zotero le récupère et le sauvegarde ;
- Si la page web contient plusieurs références (par exemple une page de résultats), alors l'icône est un dossier et il est possible de sélectionner les résultats que vous souhaitez conserver ;
- Zotero peut ajouter une référence automatiquement grâce à un identifiant ISBN ou DOI: cliquez sur « Ajouter un élément par son identifiant » dans la barre d'outils Zotero, puis tapez le numéro d'identification, l'élément sera ajouté à votre bibliothèque.

Créer

Vous pouvez aussi créer manuellement des références en sélectionnant le type de document que vous souhaitez référencer et en remplissant les champs de la notice.

Organiser

Collections:

- Dans colonne de gauche, « Mes collections » ou « Ma bibliothèque », il est possible de créer plusieurs collections;
- Créez une nouvelle collection en cliquant sur le bouton « Nouvelle collection » ;
- Il est possible de créer, renommer ou supprimer une collection (via un clic droit sur une collection);
- Pour organiser vos références il suffit de les déplacer d'un dossier d'une collection à une autre ;
- Les collections peuvent comprendre des sous-collections!

Tags et marqueurs :

- Les tags et les marqueurs servent à ajouter des informations de description ;
- Vous pouvez vous créer votre propre système de marqueurs et de tags pour organiser plus finement vos références ;

- Certaines références importées comportent parfois des marqueurs, vous pouvez vous en inspirer ou les supprimer;
- Une recherche dans Zotero interroge tous les champs, y compris les marqueurs et les tags!

Champ « note »:

- Il permet de saisir du texte libre lié à une référence précise ;
- Il peut y avoir plusieurs notes pour une même référence ;
- Il est possible de les visualiser rapidement, de les supprimer, et elles sont synchronisées si vous avez un compte Zotero;
- Une recherche dans Zotero interroge tous les champs, y compris les notes!

Exercice: Zotero

Exercice 1

- Créer une référence depuis Zotero : un livre de votre choix (titre, auteur-trice, maison d'édition, lieu, date, langue, ISBN);
- Intégrer 4 références : un article, un livre, un chapitre et une thèse ;
- Déplacer toutes vos références dans une sous-collection "HNU2000".

Exercice 2 : Traitement de texte

- Citations et bibliographies dynamiques dans des outils de traitement de texte (Word ou LibreOffice):
 - Ouvrez LibreOffice/Word et vérifiez que vous disposez bien des options de Zotero (dans la partie supérieure gauche);
 - si vous n'avez pas l'extension Zotero pour LibreOffice, installez là via le logiciel Zotero : Édition
 > Préférences > Citer > Traitements de texte > Installer le module LibreOffice
 - vous pouvez ajouter une citation d'une référence préalablement enregistrée dans Zotero;
 - une fois des citations ajoutées, vous pouvez créer une bibliographie de ces citations ;
 - si vous modifier vos références dans Zotero, vous pouvez mettre à jour votre document LibreOffice.

BibTeX

BibTeX est un format de fichier structurant une bibliographie :

- Format qui a été créé en 1985 pour gérer les bases de données bibliographiques dans des fichier LaTeX;
- Le format BibTeX fonctionne par entrée bibliographiques : chaque entrée correspond à une description d'un document avec des items « mot-clef = valeur ».

Fichier ref. bib contenant autant d'entrées que de références :

```
@article{merzeauEditorialisationCollaborativeDun2013a,
    title = {Éditorialisation collaborative d'un événement :{L}'exemple des {Entretiens} du
    volume = \{43\},
    issn = \{1168-5549\},
    shorttitle = {Éditorialisation collaborative d'un événement},
    url = {https://shs.cairn.info/revue-communication-et-organisation-2013-1-page-105},
    doi = {10.4000/communicationorganisation.4158},
    language = {fr},
    number = \{1\},
    urldate = \{2024-10-10\},\
    journal = {Communication \& Organisation},
    author = {Merzeau, Louise},
    year = \{2013\},\
   note = {Publisher: Presses universitaires de Bordeaux
Section: Info, Communication},
    pages = \{105--122\},
}
```

Better BibTeX

Better BibTex (https://retorque.re/zotero-better-bibtex/installation/) ajoute à Zotero des fonctionnalités très pratiques :

- Affichage des clés de citation (identifiants BibTeX uniques et stables);
- Nouveaux formats d'export;
- Exports dynamiques de bibliographies (synchronisation automatique des fichiers . bib).

Installez Better BibTeX

https://retorque.re/zotero-better-bibtex/installation/

Exercices

- Créez une collection avec plusieurs références que vous exportez au format BetterBibTeX avec l'option « Garder à jour » ;
- Ajouter une référence à cette collection ;
- Est-ce que le fichier BibTeX a été modifié ? ;
- Verrouillez une clé de citation d'une référence de ce export.

[Séance 5 : 07/10] REPRÉSENTER le texte (I): formats légers

Retour sur les lectures

Lectures obligatoires

• Vitali Rosati et al. (2020)

Nous le ferons dans la deuxième partie du cours.

Objectifs de la séance

Théorie:

- Comment représenter, sous forme numérique, un texte ?
- Les avantages des formats légers ?
- Principes du balisage.

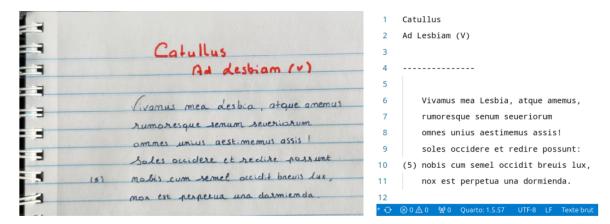
Pratique:

- Tuto Markdown;
- Tuto Stylo;
- Bonus : un peu de HTML.

Représenter le texte ?

- Représenter >< Montrer ?
- Le texte et son support!

Ce poème est-il le même ?



Catullus

Ad Lesbiam (V)

Vivamus mea Lesbia, atque amemus, rumoresque senum seueriorum omnes unius aestimemus assis! soles occidere et redire possunt: nobis cum semel occidit breuis lux, nox est perpetua una dormienda.

Le texte et l'ordinateur

- Une suite de symboles ;
- Qui peuvent être structurés;
 - À nous de définir cette structure!
- Dont il ne comprend pas le sens.

Comment choisir un format, une syntaxe, qui rende justice à un texte donné?

Encodage, format, balisage

- Encodage: façon dont les caractères sont traduits en nombres (binaires), où chaque *encodage* correspond à une *norme* (ASCII, UTF-8, etc). Plus d'informations ici;
- Format : organisation du contenu à l'intérieur d'un fichier (cf. séance précédente) ;
- Langage de balisage : façon dont le texte lui-même est structuré par des signes visibles.

Chaque format et langage impose une manière de penser le texte : .docx pour la mise en page (et encore...), .html pour la hiérarchie entre les éléments, .md pour la lisibilité et interopérabilité.

Une courte histoire du balisage

<chapter><title>Introduction to SGML</title>
<section><title>The SGML Declaration</title>
<subsection>

SGML, HTML et XML

• SGML : décrire avec des balises ;

• HTML : structurer pour afficher ;

• XML : l'arbre sémantique.

L'histoire des langages de balisage est une histoire longue, passionnante et complexe. Sans rentrer dans tous les détails, nous pouvons observer plusieurs évolutions et divergences intéressantes.

SGML est d'abord conçu pour résoudre un problème lié la représentation et la reproduction des documents. Jusqu'ici basé sur le langage Script pensé pour l'impression, les formats de document ne sont pas structuré (à part quelques exceptions comme TeX). SGML (précédé par GML), pour Standard Generalized Markup Language, introduit donc le balisage déclaratif permettant de générer des documents graphiques.

HTML est une application de SGML, plus simple et plus adapté à un affichage graphique qui ne dépend pas du système de la page imprimée. Il y a plusieurs versions successives de HTML, il faut retenir qu'il s'agit d'un ensemble d'éléments qui prennent en compte l'hypertexte. Le langage HTML est accompagné de HTTP et du système d'adresses web (les URLs).

Enfin XML est développé à partir des années 1990 dans un objectif de rendre plus générique un langage de balisage. XML est une suite de spécifications qui peuvent ensuite être implémentées dans des schémas, comme la TEI qui a d'ailleurs précédé XML. XML est puissant mais aussi complexe à utiliser, il est aujourd'hui très éloigné de la version actuelle de HTML. Ou disons que les choix techniques de HTML fait que ce langage est très permissif, ne respectant pas certaines prérogatives de SGML ou de XML.

Les formats légers

Le qualificatif « léger » permet de comprendre de quoi il s'agit face aux autres langages précédemment cités qui peuvent être jugés lourds. Grâce à une économie de caractères ou de signes pour baliser le texte, certains langages de balisage sont humainement lisibles (contrairement à HTML ou XML qui sont très verbeux). Avec quelques signes typographiques, des langages comme Markdown ou AsciiDoc permettent de faire de la sémantique avec un apprentissage minime.

Définition

Un format léger est un format :

- Textuel;
- Lisible à l'œil nu, directement dans le code (sans besoin de passer par un logiciel spécialisé);
- Non verbeux (utilise peu de signes pour représenter la structure du texte).

La structure accompagne le texte, qui reste l'élément premier.

Exemples

- .txt texte brut, sans balisage;
- .md Markdown: un format minimaliste pour structurer le texte;
- .rst reStructuredText, souvent utilisé dans la documentation Python ;
- .html parfois considéré comme "semi-léger", car il est textuel mais plus verbeux.

Une philosophie éditoriale

Trois principes:

- Lisibilité: un humain peut lire le fichier directement, sans le "compiler" ni l'ouvrir dans un logiciel particulier;
- Simplicité : la syntaxe est minimale ; on n'a pas besoin de connaître la structure interne du document ;
- Convertibilité : le format léger sert souvent de point de départ vers d'autres formats (Word, HTML, PDF, LaTeX) grâce à des outils comme Pandoc ou Quarto.

Le format texte

Un fichier au format texte est un fichier dont le code binaire (suite de 0 et de 1) se traduit par des caractères textuels uniquement (par opposition à des pixels ou des sons par exemple). Dit de manière encore plus courte : c'est un fichier qui ne contient que des caractères. (Arthur Perret, Format texte, https://www.arthurperret.fr/cours/format-texte.html)

Le terme *plein text*, largement utilisé, est une mauvaise traduction du terme anglais *plain text*. On préférera l'appellation *texte brut* en français.

Caractéristiques:

- Fichier texte non formaté qui ne nécessite aucun programme spécial pour être créé et ouvert ;
- Format extrêment simple;
- Stockage d'information en texte brut ;
- Pas de mise en forme, pas d'images, pas d'éléments multimédia ;

• Utile pour documents simples, prise de notes,...

Quelques exemples d'utilisation sur le site The Plain Text Project

Avantages:

- la **lisibilité** : contrairement aux formats binaires qui ne sont lisibles que grâce à un logiciel spécifique (comme le format .doc, et non le format .docx);
- la **légèreté** : les fichiers au format texte sont souvent très légers, ne dépassant pas quelques kilo-octets (car ne comporte que du texte). De là : rapide à lire par un programme, peu de place sur un support de stockage, utilisable par n'importe quel système d'exploitation ;
- la **pérennité** : un fichier informatique au format texte sera toujours lisible dans 10 ans, quand bien même les applications dédiées à son interprétation ne seront plus disponibles ;
- et dans une certaine mesure l'**interopérabilité** : un format texte est dépendant d'une spécification qu'il est possible d'implémenter ;
- enfin, le format texte est libre : pour l'ouvrir, le lire, le modifier ou le transmettre il n'est pas nécessaire de payer une licence.

.md

- Minimise l'impact des instructions de balisage dans le document, en insistant sur la lisibilité du texte par un être humain.
- Le document peut ensuite être traité par un algorithme pour un produire un export (HTML, PDF, DOCX, etc.)

Plus d'informations sur le blog du créateur de .md, John Gruber.

Markdown: exercice

Réalisez ce tuto Markdown

Stylo: mise en pratique

C'est quoi ?

- Éditeur de texte sémantique pour la rédaction scientifique ;
- Créé à l'UdeM par Marcello Vitali-Rosati et son équipe ;
- Destiné aux chercheur-e-s en sciences humaines ;
- Accessible en ligne uniquement (après création d'un compte);
- Pour écrire, structurer, prévisualiser, réviser, exporter (pdf, html, xml, odt, tei, etc.);
- Utilisation d'un langage de balisage léger (Markdown), d'un langage de description de métadonnées (YAML) et d'un langage de description bibliographique (BibTeX);

• Utilisation de briques logicielles existantes (et libres).

Le grand intérêt de Stylo est de permettre aux auteurs de saisir facilement leur texte dans un éditeur de texte. Ils en deviennent donc à la fois les auteurs et les encodeurs. Les concepteurs de Stylo ont souhaité construire un outil *open source* capable de garantir la bonne structuration des textes, et donc leur pérennité, leur indexation efficace, etc., sans demander aux auteurs de développer des compétences informatiques trop complexes. (Debouy, 2025, p. 79)

La documentation et des tutos en vidéo.

Discussion

• Discussion autour de (Vitali Rosati et al., 2020)

Exercice 1

- 1. Création d'un compte stylo : https://stylo.huma-num.fr;
- 2. Reprenez le texte de Catulle vu plus haut, il se trouve aussi ici ;
- 3. Ajoutez-le à votre document Stylo ; ajoutez-lui un titre, éventuellement des titres de deuxième niveau, mettez des passages en italique, cherchez-lui une traduction et ajoutez des métadonnées ;
- 4. Partagez le document avec moi : (mathilde.verstraete@umontreal.ca);
- 5. Testez la preview et faites différents exports (.html, .pdf, .docx), comparez les résultats obtenus.

Alors Stylo, WYSIWYM ou WYSIWYG?

Exercice 2

- Créez un nouvel article :
 - Idéalement, ajoutez un travail réalisé dans le cadre d'un autre cours
 - Si vous n'en avez pas, ajoutez n'importe quel article comprenant au moins :
 - * deux niveaux de titre,
 - * des liens,
 - * une citation,
 - * plusieurs paragraphes,
 - * une référence bibliographique
- Envoyez-moi le lien de prévisualisation lorsque vous avez terminé.

HTML

- HyperText Markup Language;
- Inventé au tout début des années 1990 par Tim Berners-Lee, dans le même contexte que la création du World Wide Web;
- L'idée de départ : permettre à des chercheurs de publier et relier des documents scientifiques en ligne, grâce à des liens hypertextes ;
- Premier format de publication du Web : toutes les pages web, sans exception, reposent sur du HTML (cf. ctrl-u);
- HTML sert à décrire la structure et le sens du contenu d'une page, pas sa mise en forme.

Exercice HTML

Nous allons essayer de comprendre la structure minimale d'une page html, rendez-vous ici https://codepen.io/pen

Template de départ (à copier-coller) :

Ex. 1 : Structure minimale & typographie

Dans <body>, créez :

- un en-tête avec un titre principal <h1> et un sous-titre <h2>,
- deux paragraphes (un court, un long),
- un mot en emphase () et un mot en fort ().

Ex. 2: Liens, images, listes

- Ajoutez une liste à puces (trois éléments) avec ...,
- Ajoutez un lien vers un site externe qui s'ouvre dans un nouvel onglet,
- Ajoutez une image (utilise une URL publique).

Ex. 3 : Balises sémantiques & mini défi Markdown -> HTML

• Plus compliqué : à partir de ce fragment Markdown, écrivez-en l'équivalent HTML

Catullus

Ad Lesbiam (V)

Vivamus mea Lesbia, atque amemus, rumoresque senum seueriorum omnes unius aestimemus assis! soles occidere et redire possunt:

(5) nobis cum semel occidit breuis lux, nox est perpetua una dormienda.

Références

- André, J. (1990). Petites Leçons de Typographie. Éditions du jobet.
- Bermès, E. (2023). Trente ans de numérique à la BnF. *Balisages. La revue de recherche de l'Enssib*, 6. https://doi.org/10.35562/balisages.1099
- Bot, J.-M. L. (2012). Contribution à l'histoire d'un lieu commun : l'attribution à Chateaubriand de la phrase « les forêts précèdent les peuples, les déserts les suivent ». Socio-logos . Revue de l'association française de sociologie, 7. https://doi.org/10.4000/socio-logos.2634
- Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (Éds.). (2012). Humanities to Digital Humanities. In *Digital_Humanities*. The MIT Press. https://doi.org/10.7551/mitpress/9248.003.0003
- Burnard, L. (2012). Du literary and linguistic computing aux digital humanities: retour sur 40 ans de relations entre sciences humaines et informatique. In P. Mounier (Éd.), *Read/Write Book 2: Une introduction aux humanités numériques* (p. 45-58). OpenEdition Press. https://doi.org/10.4000/books.oep.242
- Burnard, L. (2015). *Qu'est-ce que la Text Encoding Initiative*? (M. Burghart, Trad.). OpenEdition Press. https://doi.org/10.4000/books.oep.1237
- Cardon, D. (2013). Dans l'esprit du PageRank: Une enquête sur l'algorithme de Google. *Réseaux*, 177(1), 63-95. https://doi.org/10.3917/res.177.0063
- Chagué, A. (2022). Intelligence Artificielle et Intelligence Collective: Des Nouveaux Eldorados Pour Rendre Les Textes Patrimoniaux plus Accessibles? [Medium]. In *Museonum*. https://hal.science/hal-03739948 v1
- Collectif. (2018). Pourquoi Zotero? [{Billet}]. In Le blog Zotero francophone. https://doi.org/10.58079/vdwp Commission d'enrichissement de la langue française. (2019). Journal Officiel Électronique Authentifié N° 0157. https://www.legifrance.gouv.fr/download/pdf?id=rtw17ZyCloLK1ylXH84-zZnRB4DS_m6KQ1OzvTQ6mNw=
- Coombs, J. H., Renear, A. H., & DeRose, S. J. (1987). Markup Systems and the Future of Scholarly Text Processing. *Communication of the ACM*, 30(11), 933-947. https://doi.org/10.1145/32206.32209
- Dacos, M. (2011). Manifeste des Digital humanities [{Billet}]. In *THATCamp Paris*. https://doi.org/10.58079/uo27
- Dacos, M., & Mounier, P. (2015). *Humanités Numériques* [Research {{Report}}]. Institut français. https://hal.science/hal-01228945
- Daniel, J. (2022). Ode à Zotero, une décennie d'usage de Zotero [{Billet}]. In *Isidore & Ganesh*. https://doi.org/10.58079/pw5v
- Debouy, E. (2025). *Vade-Mecum Informatique Pour Lettres et Sciences Humaines* (P. universitaires de Rennes, Éd.).
- Drucker, J. (2011). Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*, 005(1). https://dhq.digitalhumanities.org/vol/5/1/000091/000091.html

HNU2000-A25 Références

Fauchié, A., & Audin, Y. (2023). The Importance of Single Source Publishing in Scientific Publishing. *Digital Studies / Le Champ Numérique*, *Special DSCN Collection #9*. https://doi.org/10.16995/dscn.9655

- France Inter. (s. d.). *Le Français qui a vu naître Google*. https://www.radiofrance.fr/franceinter/podcasts/le-code-a-change/le-francais-qui-a-vu-naitre-google-4402005
- Fredriksson, S., & Sauret, N. (2019). Écrire les communs. Au-devant de l'irréversible. *Sens public*. http://sens-public.org/dossiers/1383/
- Huma-Num. (2021). ISIDORE a 10 ans! [{Billet}]. In *Le blog d'Huma-Num et des Consortiums-HN*. https://doi.org/10.58079/pr1t
- Jahjah, M. (2017). "Annoter Le Monde et Améliorer l'humanité": Les Imaginaires Matériels d'un Logiciel d'annotation Du Web. In *Ecrilecture Augmentée Dans Les Communautés Scientifiques Humanités Numériques et Construction Des Savoirs*. https://hal.science/hal-01638125
- Just What Is TeX? (2023). In TeX Users Group. http://ftp.tug.org/whatis.html.
- Larsonneur, C. (2008). La recherche Internet en lettres et langues. Editions OPHRYS.
- Mille, A. (2014). Chapitre 2. D'Internet au web. In M. Vitali-Rosati & M. E. Sinatra (Éds.), *Pratiques de l'édition numérique* (p. 31-48). Presses de l'Université de Montréal. https://doi.org/10.4000/books.pu m.315
- Ochigame, R., & Ye, K. (2021). Search Atlas: Visualizing Divergent Search Results Across Geopolitical Borders. *Proceedings of the 2021 ACM Designing Interactive Systems Conference*, 1970-1983. https://doi.org/10.1145/3461778.3462032
- Perret, A. (2022a). À quoi sert une vue graphe ? In *arthurperret.fr*. https://www.arthurperret.fr/blog/2022-02-17-analyser-synthetiser-visualiser.html
- Perret, A. (2022b). Analyser, synthétiser, visualiser: le triptyque fiche, lien, graphe. In *arthurperret.fr*. https://www.arthurperret.fr/blog/2022-02-13-a-quoi-sert-une-vue-graphe.html
- Reboul, M., & Gefen, A. (2019). Mesures et savoirs : Quelles méthodes pour l'histoire culturelle à l'heure du big data ? *Semiotica*, 2019(230), 97. https://doi.org/10.1515/sem-2018-0103
- Robichaud, L., Burgess, J., & Linteau, P.-A. (2019). Prendre Le Tournant Spatial En Histoire: Le Laboratoire d'histoire et de Patrimoine de Montréal et Le Système de Cartographie de l'histoire de Montréal. In *Histoire et Patrimoine. Pistes de Recherche et de Mise En Valeur* (p. 213-236). Presses de l'Université Laval. https://www.pulaval.com/libreacces/9782763743295.pdf
- Schmidt, B. M. (2016). Do Digital Humanists Need to Understand Algorithms? *Debates in the Digital Humanities 2016*, 546-555. https://doi.org/10.5749/j.ctt1cn6thb.51
- Schrijver, É. (2017). Culture hacker et peur du WYSIWYG. In *Back Office*. http://www.revue-backoffice.com/numeros/01-faire-avec/eric-schrijver-culture-hacker-peur-wysiwyg
- Sinatra, M. E., & Vitali-Rosati, M. (2014). Histoire des humanités numériques. In *Pratiques de l'édition nu-mérique*. Les Ateliers de [sens public]. https://www.parcoursnumeriques-pum.ca/1-pratiques/chapitre 3.html
- Smits, T., & Wevers, M. (2023). A Multimodal Turn in Digital Humanities. Using Contrastive Machine Learning Models to Explore, Enrich, and Analyze Digital Visual Historical Collections. *Digital Scholarship in the Humanities*, 38(3), 1267-1280. https://doi.org/10.1093/llc/fqad008
- Turbiau, A. (2023). Rédaction : gérer les différences de versions avec Git [{Billet}]. In *Littératures engagées*. https://doi.org/10.58079/oe11
- Underwood, T. (2018). Why an Age of Machine Learning Needs the Humanities. In *Public Books*. https://www.publicbooks.org/why-an-age-of-machine-learning-needs-the-humanities/

HNU2000-A25 Références

Vitali Rosati, M., Sauret, N., Fauchié, A., & Mellet, M. (2020). Écrire les SHS en environnement numérique. L'éditeur de texte Stylo. *Revue Intelligibilité du numérique*, 1. https://intelligibilite-numerique.numere v.com/numeros/n-1-2020/18-ecrire-les-shs-en-environnement-numerique-l-editeur-de-texte-stylo Vitali-Rosati, M. (2018). Les Chercheurs En SHS Savent-Ils Écrire ? In *The Conversation*. http://theconversation.com/les-chercheurs-en-shs-savent-ils-ecrire-93024

Contenu

Plan du cours	2
Description du cours	2
Objectifs et contenu du cours	2
Organisation du cours	3
Modalités d'évaluation	4
Examen de mi-session	4
Travail bonus: Apprenez-moi quelque chose	4
Examen final	5
Calendrier des séances et des lectures	6
Ressources utilisées dans le cadre de ce cours	6
Plagiat et fraude	6
Crédits	6
License	6
Calendrier des séances et des lectures	7
[Séance 1 : 02/09] Introduction au cours et aux Humanités Numériques	7
[Séance 2: 09/09] COMPRENDRE son environnement	8
[Séance 3 : 16/09] CHERCHER : corpus et ressources numériques	9
[Séance 4 : 23/09] ORGANISER & BIBLIOGRAPHIER : métadonnées et formats	10
[30/09]	10
[Séance 5 : 07/10] REPRÉSENTER le texte (I) : formats légers	11
[Séance 6: 14/10] REPRÉSENTER le texte (II): formats riches et normés	12
[21/10]	12
[Séance 7 : 28/10] EXAMEN	13
[Séance 8 : 04/11] ANNOTER : des <i>marginalia</i> à l'annotation numérique	14
[Séance 9: 11/11] ÉCRIRE & PUBLIER	15
[Séance 10: 18/11] NETTOYER & EXPLOITER les données textuelles	16
[Séance 11 : 25/11] ANALYSER : textométrie, stylométrie et TAL	17
[Séance 12: 02/12] VISUALISER: vers une autre lecture	18
[Séance 13:09/12] IA	19
	20
Introduction au cours	21
Bienvenue!	
	21

HNU2000-A25 Contenu

Déroulement du cours	21
Objectifs du cours	21
Présentation des séances	22
Modalités d'évaluation	22
Séance 1 : 02/09] Introduction aux Humanités Numériques	23
Définitions	23
Humanités	23
Numérique(s)	24
Les trois moments des Humanités Numériques	24
env. 1970 : Literary and Linguistic Computing	24
env. 1980: L'humanities computing	24
env. 1990 : Digital Humanities	24
Quelques repères historiques	25
Les prémices mécaniques	25
Les concepts fondateurs	27
Vers l'ordinateur moderne	28
L'informatique et les sciences humaines	28
Démocratisation et réseau	28
Vers les humanités numériques contemporaines	30
[Séance 2 : 09/09] COMPRENDRE son environnement	31
Retour sur la lecture	31
Terminal, ligne de commande	31
La ligne de commande	31
Exemple	32
La ligne de commande, plus rapide et plus sûre ?	32
Testons	32
Où sont les fichiers sur un ordinateur ?	32
Arborescence	33
Nommer ses fichiers	33
Chemins relatifs et chemins absolus	34
Les premières commandes	34
Où suis-je?	34
Déplacer un fichier	35
Naviguer dans ses fichiers	35
Lister des fichiers	35
Créer un dossier	35
Créer un fichier	35
Supprimer un fichier	36
Afficher le contenu d'un fichier	36
Comprendre une commande	36
Quelques réflexes	

HNU2000-A25 Contenu

Exercice	36
Quelques commandes avancées	36
Bash	37
Exercices	37
WYSIWYM >< WYSIWYG?	39
WYSIWYM	40
Mettre en forme =/= structurer	40
VSCodium: installation et manipulations	40
C'est quoi Codium ?	40
Extensions	41
[Séance 3 : 16/09] CHERCHER: corpus et ressources numériques	42
Retour sur les lectures et écoutes	42
Une histoire d'internet et du Web	42
Origines	42
Principes: TCP/IP	42
Expérimentation	43
Le Web	43
Exercice	44
Les navigateurs et moteurs de recherche	44
Le Web, au commencent étaient les annuaires	44
Explorer, indexer, classer	45
Quelques moteurs de recherche	45
Google et le PageRank	47
Les évolutions des moteurs de recherche	47
Vade-Mecum d'une recherche efficace	48
Exercice	50
Une alternative aux moteurs de recherche : les flux RSS	50
[Séance 4 : 23/09] ORGANISER & BIBLIOGRAPHIER: métadonnées et formats	51
Retour sur les lectures	51
Les métadonnées	51
Pourquoi les métadonnées ?	51
Les différents formats textuels	51
Les formats : un peu d'étymologie	51
Vers une définition du format	52
Format et édition savante	52
Format et extension	52
Formats libres, formats propriétaires	53
Le format n'est pas neutre!	53
Zoom sur le libre	53
Qu'est-ce qu'une bibliographie ?	54
L'importance de structurer ses références bibliographiques	54

HNU2000-A25 Contenu

Style de citation vs style bibliographique	54
Exemple	55
Comment créer une bibliographie ?	55
Zotero	55
Qu'est-ce que Zotero?	55
Fonctionnalités	56
Un logiciel libre	56
Utilisation générale	56
Fonctionnalités proposées avec un compte	56
À quoi sert un groupe ?	57 57
Installation	57 57
De nombreuses bases de données compatibles	57 57
Description d'une utilisation classique	57 58
Quelques points importants	58
Créer	58
Organiser	58
Exercice: Zotero	59
Exercice 1	59
Exercice 2: Traitement de texte	59
BibTeX	59
Better BibTeX	60
Installez Better Bib'TeX	60
Exercices	60
Séance 5 : 07/10] REPRÉSENTER le texte (I): formats légers	61
Retour sur les lectures	61
Objectifs de la séance	61
Représenter le texte?	61
Ce poème est-il le même ?	62
Le texte et l'ordinateur	62
Encodage, format, balisage	62
Une courte histoire du balisage	63
SGML, HTML et XML	63
Les formats légers	63
Définition	64
Exemples	64
Une philosophie éditoriale	64
Le format texte	64
.md	65
Markdown: exercice	65
Stylo: mise en pratique	65
C'est quoi?	65

Contenu

Discussion		 (
Exercice 1		 (
Exercice 2		 (
HTML		 (
Exercice HTML		 (
Ex. 1 : Structure minimale & typographie		 (
Ex. 2 : Liens, images, listes		 (
Ex. 3 : Balises sémantiques & mini défi Markdown→HTML		 (
2464		c
Références		6