HNU2000-A25

Humanités Numériques : technologies

Mathilde Verstraete

Automne 2024

Table of contents

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* : À confirmer

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 and 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 ***HU2000 – 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;
- La programmation (introduction);
- etc.

Que sont les humanités numériques et quelles technologies sont utilisées dans cette approche scientifique ? 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 HNU 2000 « 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 HNU 1000 « 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 des 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.

Il est impératif que l'étudiant·e possède un ordinateur.

Calendrier des séances et des lectures

Le cours de cette année sera axé sur le cycle de vie du texte savant en contexte numérique. On abordera le texte source (séances 1-7), son analyse (8-10) et sa publication et diffusion (11-13).

[Séance 1 : 02/09] Introduction

- Présentation de l'enseignante
- Présentation des étudiant·e·s
- Présentation du cours : déroulement, courte présentation des séances, participation, évaluations ;
- Introduction aux Humanités Numériques

Lectures obligatoires

- (Sinatra and Vitali-Rosati 2014)
- (Underwood 2018)

Outils

- https://whatisdigitalhumanities.com/

Lectures supplémentaires

- (Dacos and Mounier 2015)
- (Dacos 2011)
- (Burdick et al. 2012)

[Séance 2 : 09/09] CHERCHER: corpus et ressources numériques

Lecture obligatoire	Lecture optionnelle	Théorie	Outil
TBD	TBD	Bibliothèques numériques, archives ouvertes, écosystèmes des ressources numériques	Moteurs de recherche (+regex?)

Lectures supplémentaires

- (France Inter n.d.) (podcast)
- (Cardon 2013)

[Séance 3 : 16/09] ORGANISER: métadonnées et formats

> **Note**
>
> [17/09] Date limite pour la modification des choix de cours

Lecture obligatoire	Lecture optionnelle	Théorie	Outil
TBD	TBD	Les différents formats textuelsLes traitements *vs* éditeurs de texte	Pandoc

[Séance 4 : 23/09] ANNOTER: des *marginalia* à l'annotation numérique

Lecture obligatoire	Lecture optionnelle	Théorie	Outil
TBD	TBD	Typologie de l'annotation, enjeux collaboratifs, question de la trace	Hypothes.is

[30/09]

> **Note**
>
> Congé universitaire – Journée nationale de la vérité et de la réconciliation

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

Lecture obligatoire	Lecture optionnelle	Théorie	Outil
J. H. Coombs, Allen H Renear, et Steven J DeRose. « Markup Systems and the Future of Scholarly Text Processing »	TBD	Introduction à .txt, .md, .html	Prise en main de MarkdownStylo

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

Lecture obligatoire	Lecture optionnelle	Théorie	Outil
LaTeX + Introduction à la TEI	TBD	Baliser son texte	Prise en main de LaTeX (Over-Leaf)Observation d'un document en XML-TEI (LeafWriter)Interopérabilité des formats (ekdosis)

[21/10]

> **Note**
>
> Semaine de lecture – pas de cours

[Séance 7 : 28/10] EXAMEN

> **!** Important
>
> Examen de mi-session

[Séance 8 : 04/11] BIBLIOGRAPHIER: citations, identifiants, interopérabilité

> **i** Note
>
> [07/11] Date limite pour l'abandon d'un cours

Lecture obligatoire	Lecture optionnelle	Théorie	Outil
TBD	TBD	BibTeX, CSL	Zotero

[Séance 9 : 11/11] NETTOYER les données textuelles

Lecture obligatoire	Lecture optionnelle	Théorie	Outil
TBD (data captcha?)	TBD	OCR, HTR, dirty OCR	Introduction à eScriptorium (?)

[Séance 10 : 18/11] ANALYSER: textométrie, stylométrie et TAL

Lecture obligatoire	Lecture optionnelle	Théorie	Outil
TBD	TBD	Ce que nous disent les mots	Voyant + Notebook Python

[Séance 11 : 25/11] VISUALIER: vers une autre lecture

Lecture obligatoire	Lecture optionnelle	Théorie	Outil
TBD	D. Keim et al., « Visual Analytics: Definition, Process, and Challenges »Drucker, Humanities Approaches to Graphical Display	Les différents types de visualisation	TimeLineJS, StoryMapJS

[Séance 12 : 02/12] ÉCRIRE & PUBLIER

Lecture obligatoire	Lecture optionnelle	Théorie	Outil
TBD	Fauchié, The Importance of Single Source Publishing in Scientific Publishing	Le SSP et les chaînes de publication + introduction à Git	Mise en place d'un workflow éditorial

[Séance 13 : 09/12] IA

Lecture obligatoire	Lecture optionnelle	Théorie	Outil
TBD	TBD	Définition, panorama (algorithmes, modèles de langage, IA générative)	Discussion?

[Séance 14 : 16/12]

> **!** Important
>
> Examen final

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%;
- Examen de mi-session (théorie) : 25%
- Examen final (théorie) : 25%
- Travail final (pratique) : 25%

L'examen est composé de deux parties :

1. deux examens écrits, qui consistent en un QCM et des questions ouvertes portant sur la théorie vue en cours et dans les lectures ;
2. un travail portant sur la rédaction d'un projet en utilisant au moins deux des outils vus en cours.

Ressources utilisées dans le cadre de ce cours

- Le présent site 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 »

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

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

1 [Séance 1 : 02/09] Introduction

1.1 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.

1.2 Tour de table

Quelques questions :

- quel est votre nom ?
- quel est votre parcours académique ?
- qu'attendez-vous de ce cours ?
- que savez-vous des Humanités Numériques ?

1.3 Déroulement du cours

- Cours en présentiel = présence requise ;
- Lectures obligatoires avant chaque séance ;
- Support/ressources disponibles à la fin de chaque cours (site web dédié) ;
- Pendant les séances : partie théorique, manipulations, échanges ;
- Exposés à prévoir sur certaines séances.

Pour chaque séance vous devez lire un texte, souvent de quelques pages. Ce texte servira de base à une discussion pendant la séance, vous aurez par ailleurs besoin de les connaître pour les différents travaux à réaliser.

Les supports de cours, y compris une partie de mes notes, seront placés à la suite de chaque séance sur un site web dédié. Vous pourrez donc les consulter à tout moment. Attention, cela ne vous dispense pas de prendre des notes pendant le cours !

1.4 Objectifs du cours

Cf. le Plan de cours

1.5 Présentation des séances

Cf. le Plan de cours

1.6 Modalités d'évaluation

Cf. le Plan de cours

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

2.1 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 and 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.[1](Commission d'enrichissement de la langue française 2019)

- *Quelques* autres définitions ici ou ici.

2.2 Humanités...

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

[1]Cette définition a été réutilisée sur le site de l'OQLF.

2.3 …Numérique(s)

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

2.4 Les trois moments des Humanités Numériques

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

2.4.1 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

2.4.2 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*)

2.4.3 env. 1990 : *Digital Humanities*

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

2.5 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, 17)

2.5.1 Les prémices mécaniques

- II^{e} 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'engrenage. 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 2.1: 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.

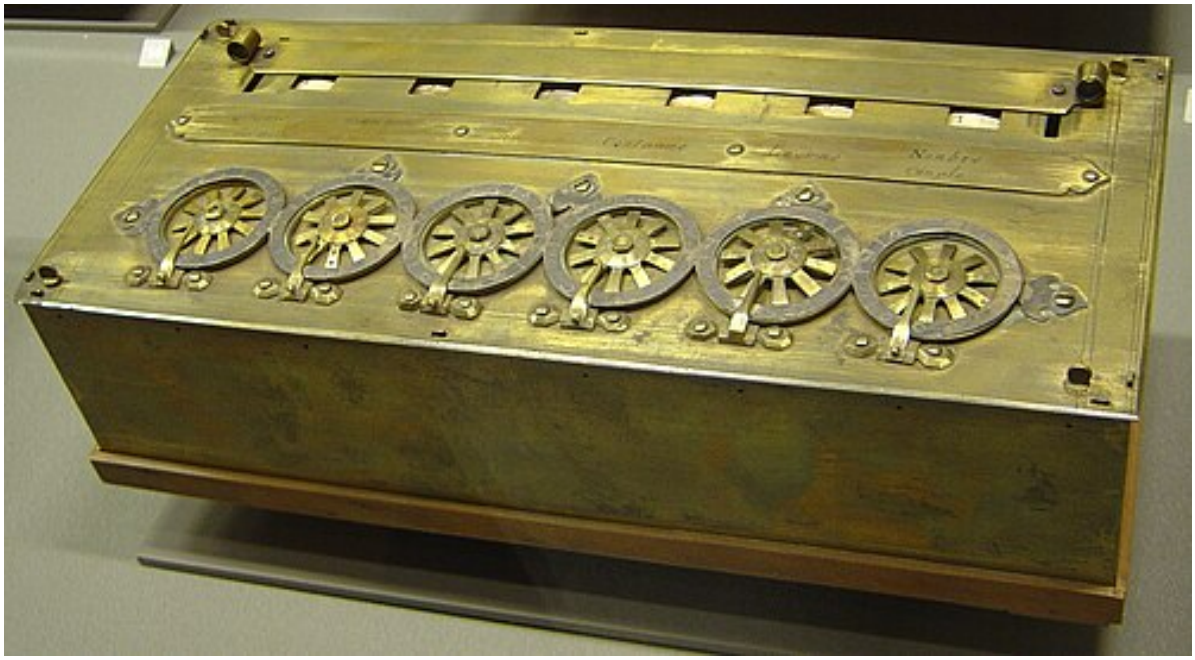

Figure 2.2: Pascaline, conservée au musée des Arts et Métiers, Paris

- 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.

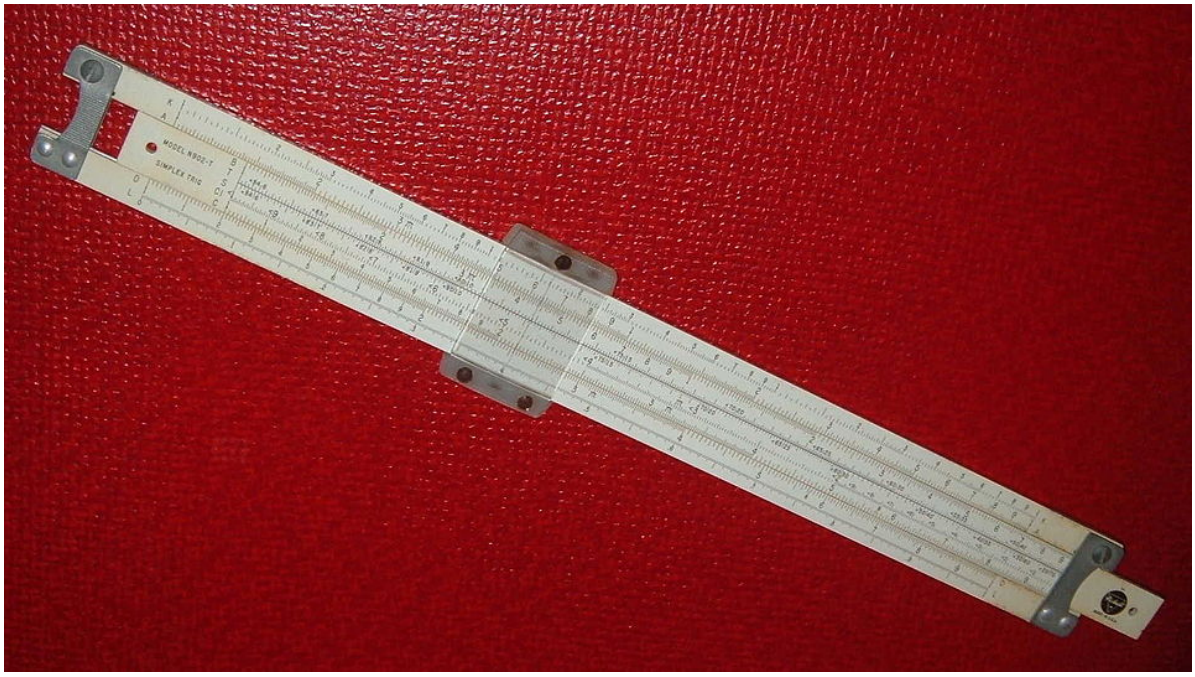

Figure 2.3: Règle à calcul scolaire, 10 pouces (Pickett N902-T simplex trig).

- 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.

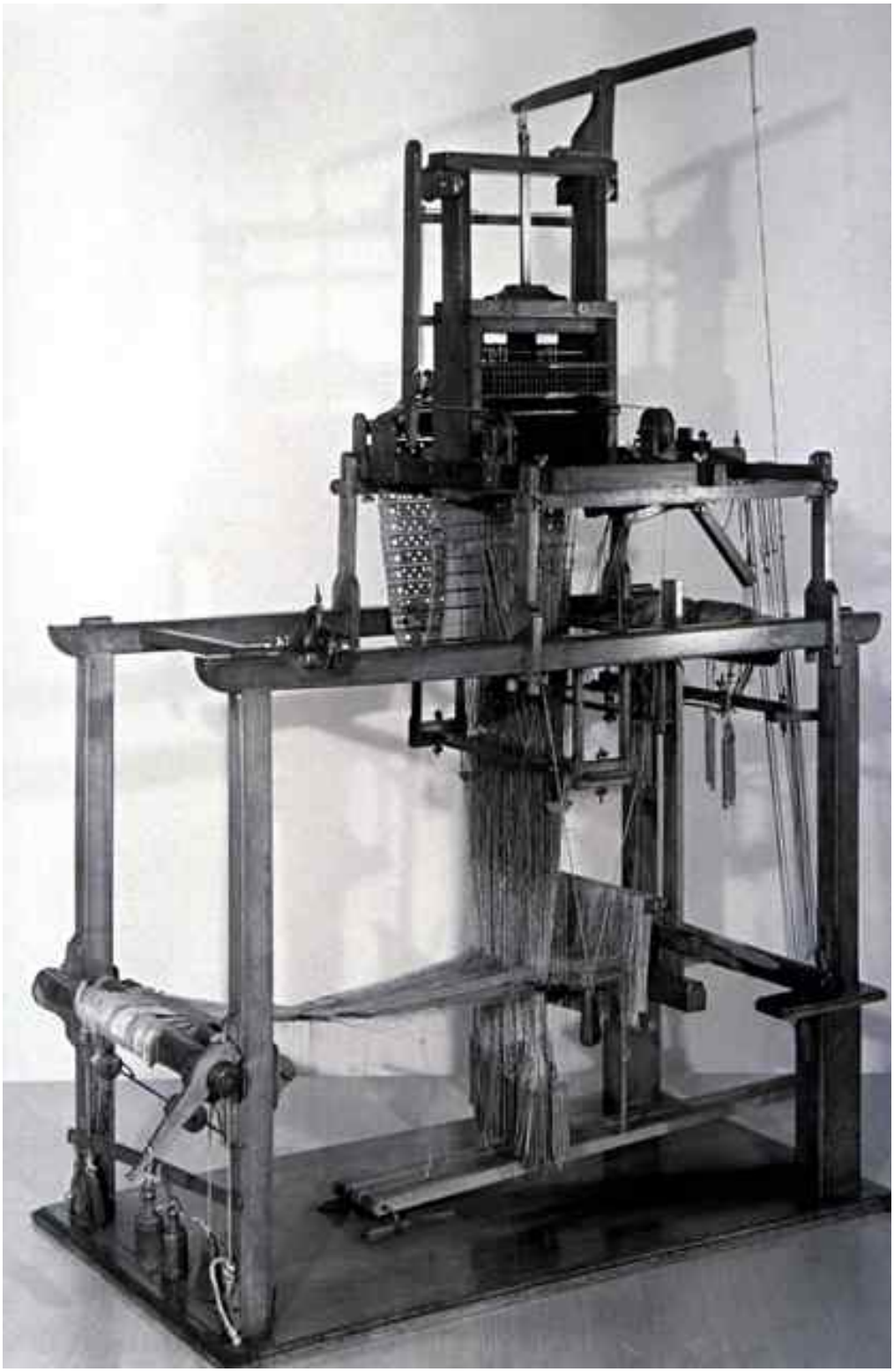

Figure 2.4: Métier à tisser de J.-M. Jacquard

2.5.1.1 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.

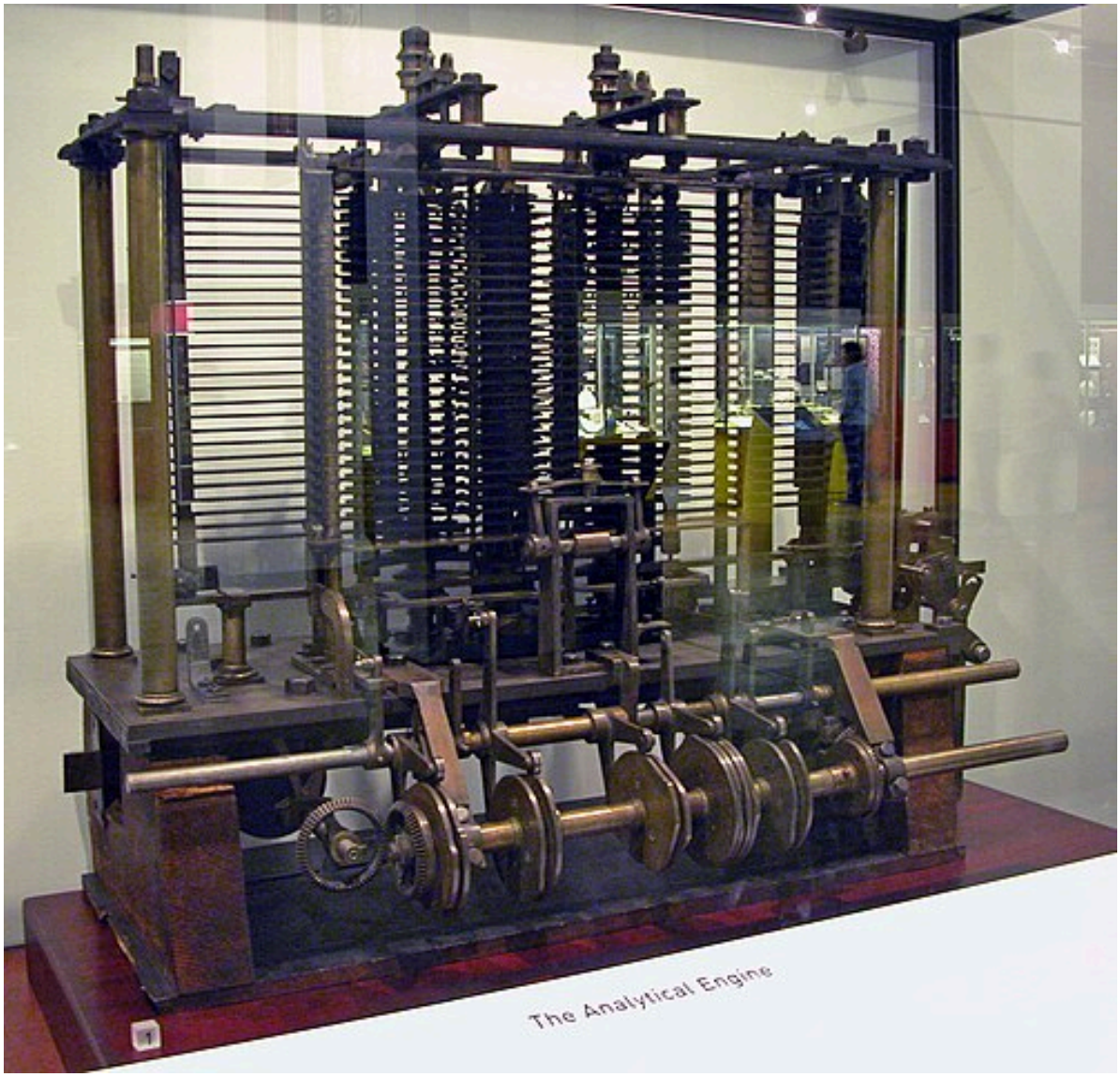

Figure 2.5: Prototype (1871) non terminé de la machine analytique de Babbage

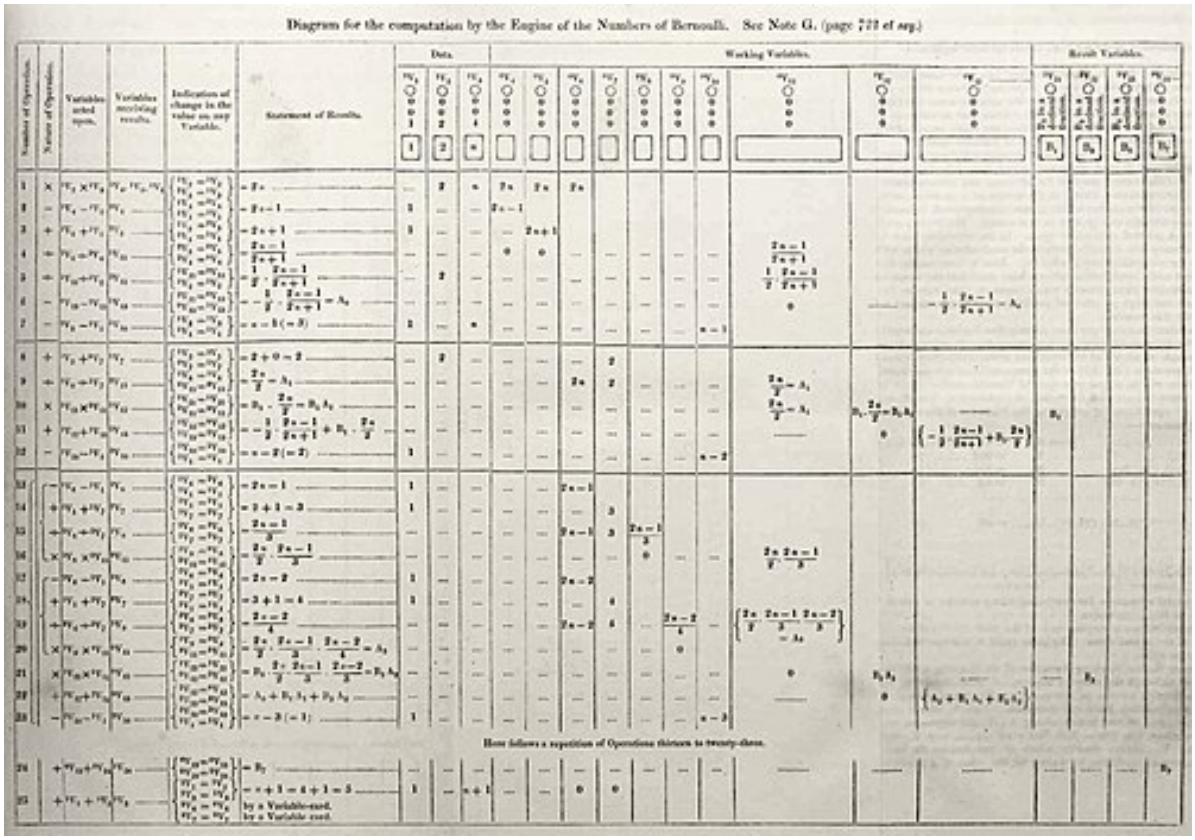

Figure 2.6: « Programme » de calcul des nombres de Bernoulli dans une note d'A. Lovelace (1843)

- 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*)[2].

2.5.1.2 Vers l'ordinateur moderne

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

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

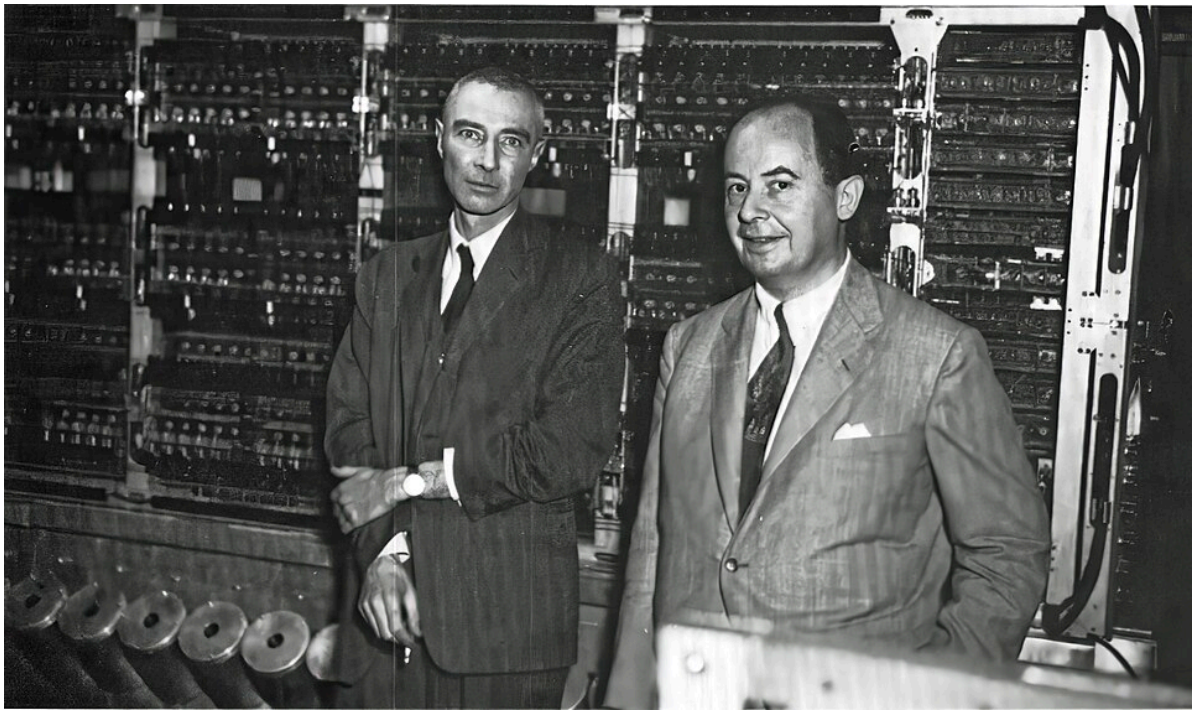

Figure 2.7: Oppenheimer et von Neumann à l'inauguration en 1952 de l'ordinateur construit pour l'*Institute for Advanced Study*.

- Années 1950 : IBM commercialise ses premiers ordinateurs ; FORTRAN[3] (FORmula TRANslator) devient le premier langage de haut niveau utilisé massivement pour le calcul scientifique et l'ingénierie.

2.5.1.3 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.

2.5.1.4 Démocratisation et réseau

- Années 1960–1970 : naissance d'Internet, à la suite de projets militaires et universitaires (comme ARPANET, 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.

[3]Un 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 2.8: IBM PC 5150 avec clavier

- 1990 : Tim Berners-Lee invente le World Wide Web au CERN, rendant possible la circulation mondiale des textes numériques. Les concepts d'Internet et du Web sont complètement différents ! Le Web repose sur Internet, mais il en est une application.

Figure 2.9: Tim Berners-Lee

- 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).

2.5.1.5 Vers les humanités numériques contemporaines

- Années 2000-2010 : massification du numérique en SHS : OCR3HTR, 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.

3 [Séance 2 : 09/09] CHERCHER: corpus et ressources numériques

3.1 Une histoire d'internet et du Web

Nous avons évoqué, la semaine passée, Internet 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)

- Dimension physique de l'infrastructure d'Internet qui permet la possibilité de l'existence du Web. Voir le documentaire *World Brain* par Stéphane Degoutin et Gwenola Wagon, visible ici.

3.1.1 Origines

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

Plus de détails dans (Mille 2014).

3.1.2 Principes : TCP/IP

TCP/IP : TCP (Transmission Control Protocol) et 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 (c'est ce qui permet à Internet d'être un réseau fiable) ;
- application : il s'agit du Web, mais aussi d'autres applications comme le partage de fichiers (FTP) ou le courriel (IMAP et SMTP).

> **Tip**
>
> Pour en savoir plus sur Internet et ses protocoles, voici une série de vidéos très complètes : https://iletaitunefoisinternet.fr/.

3.1.3 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

3.1.4 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.

3.1.5 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/`

3.2 Les 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)

3.2.1 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.

Figure 3.1: Yahoo! en 2001

3.2.2 Constituer des index

Les moteurs de recherche, apparus au milieu des années 1990 (Altavista, Lycos, Excite…), fonctionnent selon deux grandes opérations :

- Indexation : des robots (« crawlers » ou « spiders ») parcourent le Web de lien en lien et enregistrent les pages dans des bases de données.
- Recherche : lorsqu'un utilisateur formule une requête, le moteur interroge cet index, applique ses algorithmes et hiérarchise les résultats.

3.2.3 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).

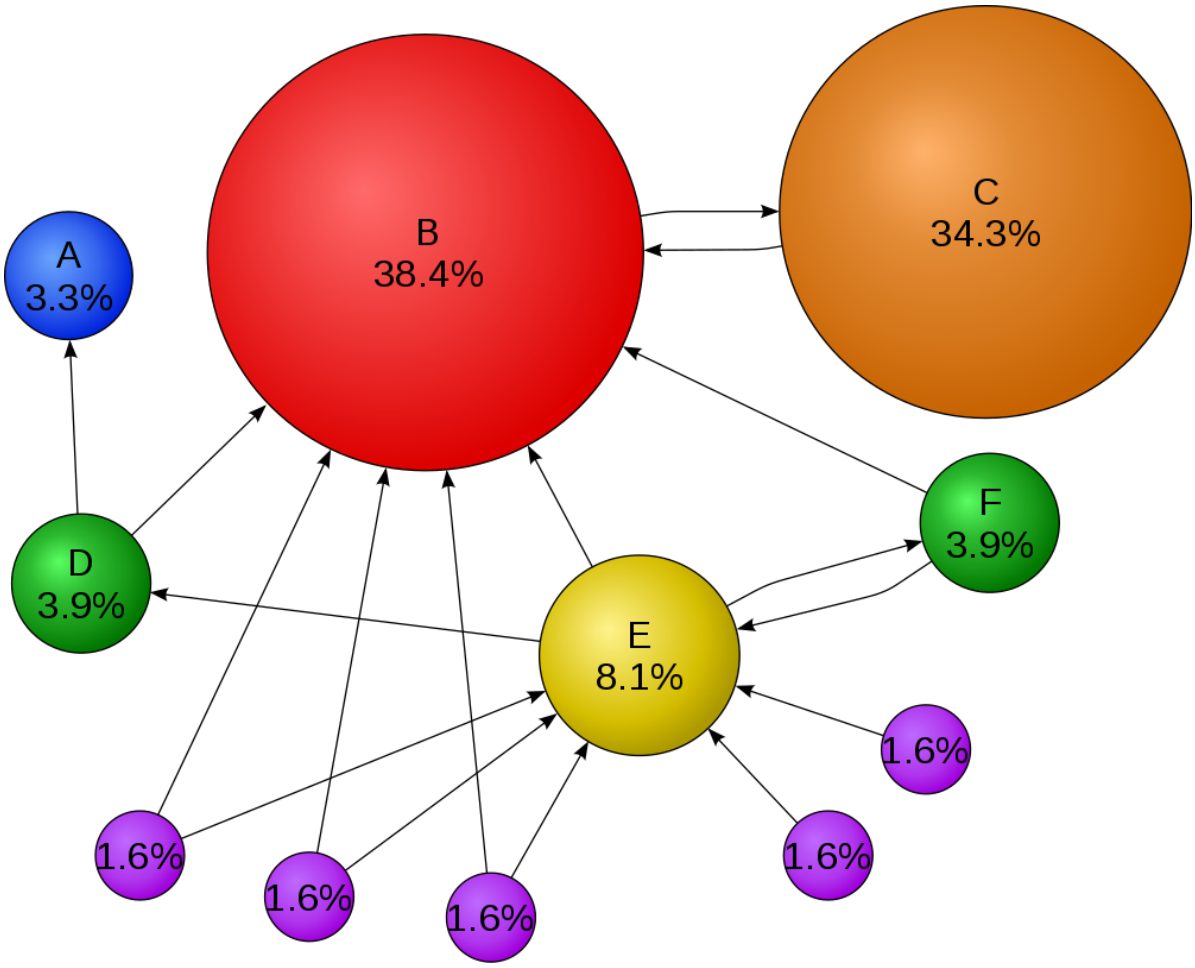

Figure 3.2: Schéma du PageRank

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).

Pour en savoir plus sur le PageRank, écouter le podcast (France Inter n.d.) (à partir de la 21e mn.) ou lisez (Cardon 2013)

3.2.4 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.

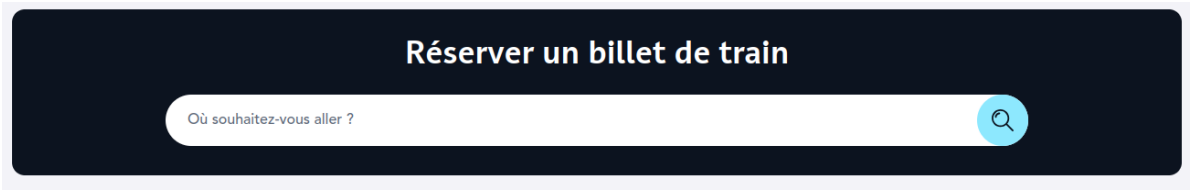

Figure 3.3: Outil de réservation SNCF, capture d'écran, 18 août 2025

Fonctionnement d'un moteur de recherche :

- Des robots parcourent le Web, de liens en liens ;

- Les données sont indexées et classées.
- L'utilisateur formule une requête.
- Le moteur sélectionne et hiérarchise les résultats grâce à des algorithmes.
- Les requêtes influencent à leur tour les robots et les algorithmes (boucle d'apprentissage).

3.2.5 Vade-Mecum d'une recherche google

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. Combiner des opérateurs

- `AND`, `OR`,`-` (exclusion) permettent de préciser la recherche. Des parenthèses servent à grouper des termes ou des opérateurs.
- e.g. : `"humanités numériques" AND "édition" -cours` → 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".

3.2.6 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

References

Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp, eds. 2012. "Humanities to Digital Humanities." In *Digital_Humanities*, 0. The MIT Press. https://doi.org/10.7551/mitpress/9248.003.0003.

Burnard, Lou. 2012. "Du literary and linguistic computing aux digital humanities : retour sur 40 ans de relations entre sciences humaines et informatique." In *Read/Write Book 2 : Une introduction aux humanités numériques*, edited by Pierre Mounier, 45–58. Read/Write Book. Marseille: OpenEdition Press. https://doi.org/10.4000/books.oep.242.

Cardon, Dominique. 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.

Commission d'enrichissement de la langue française. 2019. "Journal Officiel Électronique Authentifié n° 0157."

Dacos, Marin. 2011. "Manifeste des Digital humanities." {Billet}. *THATCamp Paris*. https://doi.org/10.58079/uo27.

Dacos, Marin, and Pierre Mounier. 2015. "Humanités Numériques." Research {{Report}}. Institut français.

Debouy, Estelle. 2025. *Vade-Mecum Informatique Pour Lettres Et Sciences Humaines*. Edited by Presses universitaires de Rennes. Didact Méthodes.

France Inter. n.d. "Le Français qui a vu naître Google." Accessed August 18, 2025.

Mille, Alain. 2014. "Chapitre 2. D'Internet au web." In *Pratiques de l'édition numérique*, edited by Marcello Vitali-Rosati and Michael E. Sinatra, 31–48. Parcours numérique. Montréal: Presses de l'Université de Montréal. https://doi.org/10.4000/books.pum.315.

Sinatra, Michaël E., and Marcello Vitali-Rosati. 2014. "Chapitre 3. Histoire des humanités numériques." In *Pratiques de l'édition numérique*, edited by Michael E. Sinatra, 49–60. Parcours numérique. Montréal: Presses de l'Université de Montréal. https://doi.org/10.4000/books.pum.317.

Underwood, Ted. 2018. "Why an Age of Machine Learning Needs the Humanities." *Public Books*.