



Semestre	Code UE	Nom UE	ECTS	Enseignements		Heures maquette	
						Enseignements HN	Enseignements disciplinaires
S1	UE 1.1	Histoire et écosystème des HN	4	Histoire et écosystème des HN		21	
	UE 1.2	Constitution de corpus	4	Traitement de documents textuels (niv 1)		21	
	UE 1.3	Informatique et statistique pour les HN	8	Algorithmique et programmation		21	
				Statistique exploratoire		21	
	UE 1.4	ALLSHS	10	Séminaires ou enseignements disciplinaires	Validé dans le master disciplinaire		63
	UE 1.5	Enseignements complémentaires du master disciplinaire	4	Langue vivante étrangère ou autre enseignement ou formation à la recherche	Validé dans le master disciplinaire		21
				Total semestre 1		84	84
S2	UE 2.1	Histoire et anthropologie du numérique	4	Histoire et anthropologie du numérique		21	
	UE 2.2	Bases de données et Web	8	Bases de données relationnelles		21	
				Technologies web		21	
	UE 2.3	Spécialisation en HN	4	Traitement de documents textuels et visuels (niv 2)	Statistique et enquête pour les SHS	21	
	UE 2.4	ALLSHS	10	Séminaires ou enseignements disciplinaires	Validé dans le master disciplinaire		63
	UE 2.5	Enseignements complémentaires du master disciplinaire	4	Langue vivante étrangère ou autre enseignement ou formation à la recherche	Validé dans le master disciplinaire		21
				Total semestre 2		84	84

Semestre	Code UE	Nom UE	ECTS	Enseignements		Heures maquette	
						Enseignements HN	Enseignements disciplinaires
S3	UE 3.1	Enjeux sociétaux des HN	4	Enjeux épistémologiques, politiques, juridiques, éthiques du numérique		21	
	UE 3.2	Analyse de données	4	Analyse de données : textes, images et réseaux		21	
	UE 3.3	Spécialisation en HN	12	Edition numérique	Data visualisation	21	
				Bases de données semi-structurées	Machine Learning pour l'analyse des données structurées et des réseaux	21	
				Modélisation et Web de données	Machine Learning pour les données textuelles	21	
	UE 3.4	ALLSHS	6	Séminaires ou enseignements disciplinaires	Validé dans le master disciplinaire		63
	UE 3.5	Enseignements complémentaires du master disciplinaire	4	Langue vivante étrangère ou autre enseignement ou formation à la recherche	Validé dans le master disciplinaire		21
			Total semestre 3		105	84	
S4	UE 4.1	Conduite de projets en HN	4	Conduite de projets en HN		21	
	UE 4.2	Stage et mémoire	6	Stage (4 mois minimum)			
			20	Mémoire			
			Total semestre 4				

Référentiel de l'offre de formation

Diplôme : Master 1 Humanités numériques

Semestre : 1

Nom de l'enseignement : Histoire et écosystème des HN

Nom du ou des enseignants : Marianne Reboul

Objectifs de l'enseignement : Découvrir le panorama des pratiques HN en Europe

Prérequis : aucun

Compétences développées : Compréhension critique de projets existants

Descriptif : Le module aura pour but de faire découvrir aux participants ce qui se fait en Humanités Numériques en Europe, et d'ouvrir des horizons pour recherches de chacun (tant scientifiques que professionnelles). Le cours abordera l'histoire des humanités numériques, en mettant en lumière leurs enjeux à travers le temps. Il permettra aussi de rencontrer différents acteurs des humanités numériques, de tous milieux professionnels. Enfin, il abordera les problématiques actuelles des Humanités Numériques, à commencer par la question du partage des données et de l'Open Data.

Logiciel et/ou matériel utilisé : aucun spécifiquement (mais les intervenants y auront sûrement recours)

Modalités des contrôles de connaissance : Rapport sur un article scientifique (coef 1), Projet de fin de semestre (coef 3).

Référentiel de l'offre de formation

Diplôme : Master 1 Humanités numériques

Semestre : 1

Nom de l'enseignement : Traitement de documents textuels (niv 1)

Nom du ou des enseignants : Bruno Bureau

Objectifs de l'enseignement : apprendre à constituer un corpus numérique à partir de documents textuels encodés et annotés, en intégrant des documents multimedia. Construire l'architecture du corpus, insérer les liens et relations nécessaires entre les éléments et produire une visualisation basique du résultat.

Prérequis : Niveau initiation aucun prérequis attendu

Compétences développées :

Convertir en numérique des objets existant sur supports traditionnels

Constituer un corpus de documents

Constituer un jeu de données (d'enquête, ouvertes, etc.)

Encoder un corpus

Annoter et indexer des documents

Décrire des ressources numériques selon les normes

Comprendre de manière critique et réflexive les données, les méthodes et les outils numériques

Identifier, sélectionner et analyser avec esprit critique diverses ressources spécialisées pour documenter un sujet et synthétiser ces données en vue de leur exploitation

Descriptif : voir ci-dessus

Logiciel et/ou matériel utilisé : Oxygen XML editor

Modalités des contrôles de connaissance :

QCM sur des questions de cours, et examen sur machine (1h) mobilisant les savoir-faire acquis pendant le cours.

Référentiel de l'offre de formation

Diplôme : Master 1 Humanités numériques

Semestre : 1

Nom de l'enseignement : Algorithmique et programmation

Nom du ou des enseignants : Jérôme Darmont

Objectifs de l'enseignement :

- Introduction au raisonnement algorithmique
- Pratique d'un langage de programmation (Python)

Prérequis : Niveau C2i ou PIX

Compétences développées :

- Connaître les principes de base de l'algorithmique et de la programmation
- Programmer des algorithmes simples (boucles, tests, sous-programmes, listes)
- Manipuler des fichiers textes
- Utiliser des modules préfabriqués

Descriptif :

0. Introduction
1. Stocker et manipuler des données en mémoire (variables, constantes, entrées/sorties)
2. Structures alternatives (tests)
3. Structures itératives (boucles)
4. Sous-programmes (fonctions, procédures, modules)
5. Structures de données avancées (chaînes de caractères, listes)
6. Stocker des données persistantes (fichiers textes)

Logiciel et/ou matériel utilisé : Python/IDLE

Modalités des contrôles de connaissance :

- Rendu de TD -- Coef. 25 %
- Épreuve sur machine (2h, documents autorisés) -- Coef. 75 %

Référentiel de l'offre de formation

Diplôme : Master 1 Humanités numériques

Semestre : 1

Nom de l'enseignement : Statistique exploratoire

Nom du ou des enseignants : Sabine LOUDCHER

Objectifs de l'enseignement :

- Présentation des concepts usuels de statistique uni et bivariable (vocabulaire, tableaux, graphiques, indicateurs)
- Utilisation de données ouvertes : principes, récupération, préparation/nettoyage des données
- Initiation au logiciel R pour réaliser des traitements statistiques
- Choix de la ou des méthodes statistiques adaptées au problème et aux données
- Analyse critique des méthodes utilisées et des résultats obtenus

Prérequis : aucun

Compétences développées :

- ✓ Savoir programmer pour analyser et visualiser des données
- ✓ Analyser des données grâce à des techniques de statistique
- ✓ Comprendre de manière critique et réflexive les données, les méthodes et les outils numériques

Descriptif :

Séance 1 : (Re) Présenter les données et introduction au logiciel R

Séance 2 : Résumer les données et formation au logiciel R

Séance 3 : Croiser l'information et formation au logiciel R

Séance 4 : Importance de la visualisation et découverte de RMarkdown

Séance 5 : une étude statistique complète et découverte de la cartographie

Séance 6 : Données ouvertes ou existantes (principes de l'open data ; collecte et aspiration ;
préparation/nettoyage de données ouvertes ou existantes)

Séance 7 : évaluation

Logiciel et/ou matériel utilisé :

- Logiciels R et RStudio
- Apprentissage de la statistique par la pratique grâce à l'outil informatique et à l'expérimentation
- A partir d'exemples concrets, réflexion et présentation des concepts théoriques
- Mise en pratique des savoirs sur des problèmes réels liés aux Humanités Numériques

Modalités des contrôles de connaissance :

Devoir individuel sur le cours + Cas pratique sur machine

Référentiel de l'offre de formation

Diplôme : Master 1 Humanités numériques

Semestre : 2

Nom de l'enseignement : Histoire et anthropologie du numérique

Nom du ou des enseignants : Pascal ROBERT et Agnieszka TONA

Objectifs de l'enseignement :

- Comprendre et savoir analyser l'éco-système économico-gestionnaire et politique du numérique dans une perspective historique longue (depuis la fin du XVIII^es) ;
- Comprendre des usages et des modes d'appropriation des objets informationnels dans l'univers numérique.

Prérequis : Culture générale en histoire économique, politique et sociale

Compétences développées :

- Savoir analyser de manière critique le développement économico-gestionnaire et politique du numérique ;
- Connaître dans quelle mesure les transformations liées au numérique affectent aujourd'hui la nature même du document ;
- Disposer d'un recul critique sur un certain nombre de discours entendus aujourd'hui sur les évolutions des notions concernées.

Descriptif :

La première partie du cours présente les dimensions géo-politique et économico-gestionnaire de l'histoire des télécommunications, de l'informatique et d'internet. Il montre comment ces technologies ont participé à l'instauration de deux moments de la mondialisation au XIX^e et dans la deuxième moitié du XX^es. Il souligne deux phénomènes de fond : la constante accélération du politique et de l'économico-gestionnaire que ces technologies équipent techniquement depuis deux siècles et le développement de la logique de la gestionnarisation de la société.

La deuxième partie de ce cours s'inscrit dans la continuité des thèmes abordés dans sa première partie. En s'appuyant sur l'exemple du document, elle vise à faire comprendre aux étudiants dans quelle mesure la technique, et plus particulièrement les transformations liées au numérique affectent aujourd'hui la nature même du document. Elle met l'accent sur les principaux changements et innovations « technologiques » à travers les siècles, qui ont un impact aujourd'hui, c'est-à-dire à « l'ère du numérique », sur le document. Ces différentes évolutions socio-historiques du document y sont analysées tant en termes de changements de supports et de techniques que de pratiques de lecture.

Logiciel et/ou matériel utilisé : aucun

Modalités des contrôles de connaissance : deux devoirs surveillés

Référentiel de l'offre de formation

Diplôme : Master 1 Humanités numériques

Semestre : 2

Nom de l'enseignement : Bases de données relationnelles

Nom du ou des enseignants : Agnieszka TONA

Objectifs de l'enseignement : Faire acquérir un premier socle des connaissances et compétences nécessaires pour gérer les données stockées dans des bases de données relationnelles.

Prérequis : Culture générale en informatique acquise durant le semestre 1.

Compétences développées :

- Connaître les fonctionnements des bases de données et des Systèmes de Gestion de Base de Données Relationnelles (SGBDR) et savoir les gérer et administrer ;
- Savoir formuler des requêtes en langage d'interrogation et de manipulation de bases de données : le SQL (Structured Query Language) ;
- Être capable de manipuler des données structurées selon le modèle relationnel en utilisant le langage SQL ;
- Savoir modéliser et concevoir une base de données relationnelles.

Descriptif :

Les données sont aujourd'hui omniprésentes dans nos environnements tant personnels que professionnels. Quel que soit votre futur métier vous serez inévitablement amenés à gérer (c.-à-d. à créer, collecter, mettre à jour, interroger, administrer, migrer ou détruire, etc.) les données stockées dans des bases de données ou entrepôts de données, devenus aujourd'hui la solution de référence en matière de gestion de l'information au sein des établissements publics ou entreprises. Ce cours a pour objectif de vous faire acquérir des connaissances et compétences nécessaires pour le faire.

Nous y mettrons l'accent surtout sur l'acquisition des principes des SGBDR – Systèmes de Gestion de Base de Données Relationnelles (qu'est-ce qu'une base de données, qu'un SGBDR, quelles sont leurs propriétés et fonctionnalités, quels sont leurs principaux langages d'interrogation et de manipulation de données, etc.) et sur la mise en pratique de ces principes, à travers un ensemble d'exercices pratiques avec les SGBD *Access*, *MySQL* ou *MariaDB*.

Logiciel et/ou matériel utilisé : SGBD *Access*, *MySQL*, *MariaDB*

Modalités des contrôles de connaissance : Devoir individuel + Projet transversal à la spécialité HN et aux technologies web

Référentiel de l'offre de formation

Diplôme : Master 1 Humanités numériques

Semestre : 2

Nom de l'enseignement : Technologies Web

Nom du ou des enseignants : Isabel Colon de Carvajal + Benjamin Abrial (à confirmer)

Objectifs de l'enseignement :

- Présentation du Web et Internet (architecture réseau ; client/serveur ; protocoles HTTP/FTP ; TCP ; DNS ; URL)
- Aperçu des différentes technologies web et de leurs fonctionnalités (Java, PHP, MySQL, HTML, XML, etc.)
- Maîtrise des langages HTML/CSS
- Introduction aux CMS et prise en main de Wordpress
- Crawl du web

Prérequis : Aucun prérequis

Compétences développées :

- Mise en œuvre de technologies du web
- Connaître le cycle de conception d'un site web
- Concevoir un site web avec un CMS
- Convertir en numérique des objets existant sur supports traditionnels
- Constituer un corpus de documents ou un jeu de données (d'enquête, ouvertes, etc.)
- Interroger une base de données de documents ou une base de données relationnelles

Descriptif :

- À partir d'exemples concrets, réflexion et présentation des concepts théoriques
- Mise en pratique du langage HTML/CSS à l'aide d'un éditeur de codes sources
- Mise en pratique de mini programme en Java, PHP et mise en relation avec des bases de données MySQL
- Réalisation d'un site web avec un CMS dans le cadre d'un projet transversal associé aux cours « Bases de données »

Logiciel et/ou matériel utilisé :

- Éditeur de texte : Note Pad++ ou Oxygen
- phpMyAdmin
- Filezilla
- Wordpress

Modalités des contrôles de connaissance :

- Un DS individuel
- Un site Web ancré dans un projet transversal

Référentiel de l'offre de formation

Diplôme : Master 1 Humanités numériques

Semestre : 2

Nom de l'enseignement : Statistique et enquête pour les SHS

Nom du ou des enseignants : Sabine LOUDCHER

Objectifs de l'enseignement :

- Collecte de données en SHS (par enquête, données ouvertes)
- Préparation et nettoyage des données
- Méthodes factorielles de réduction de dimension (ACP, AFC, ACM)
- Méthodes usuelles de classification non supervisée des données (clustering, CAH, k-Means)
- Choix de la ou des méthodes statistiques adaptées au problème et aux données
- Analyse critique des méthodes utilisées et des résultats obtenus

Prérequis : Statistique descriptive uni et bivariable (tableaux, graphiques et indicateurs usuels)

Compétences développées :

- ✓ Savoir formuler une problématique en HN
- ✓ Comprendre le besoin d'un projet
- ✓ Identifier, sélectionner et analyser avec esprit critique diverses ressources spécialisées pour documenter un sujet et synthétiser ces données en vue de leur exploitation
- ✓ Constituer un jeu de données (d'enquête, ouvertes, etc.)
- ✓ Savoir programmer pour analyser et visualiser des données
- ✓ Analyser des données grâce à des techniques de statistique
- ✓ Valoriser les données grâce à des techniques de visualisation
- ✓ Comprendre de manière critique et réflexive les données, les méthodes et les outils numériques

Descriptif :

Séance 1 : Collecte de données en SHS (par enquête, données ouvertes)

Séance 2 : Analyse en composantes principales (ACP)

Séance 3 : Analyse factorielles des correspondances (AFC)

Séance 4 : Analyse des correspondances multiples (ACM)

Séance 5 : Clustering des données (CAH, k-Means)

Séance 6 : Etude de cas complète

Séance 7 : Evaluation

Logiciel et/ou matériel utilisé :

- Logiciels R et RStudio
- A partir d'exemples concrets, réflexion et présentation des concepts théoriques
- Mise en pratique des savoirs sur des problèmes réels liés aux Humanités Numériques

Modalités des contrôles de connaissance :

Cas pratique sur machine + Devoir individuel sur le cours + Projet transversal aux bases de données et technologies web

Référentiel de l'offre de formation

Diplôme : Master 1 Humanités numériques

Semestre : 2

Nom de l'enseignement : Traitement de documents textuels (niv 2)

Nom du ou des enseignants : Bruno Bureau, Marianne Reboul

Objectifs de l'enseignement : apprendre à constituer un corpus numérique à partir de documents textuels encodés et annotés, en intégrant des documents multimedia. Construire l'architecture du corpus, insérer les liens et relations nécessaires entre les éléments et produire une visualisation basique du résultat. Découverte d'outils de numérisation et d'acquisition de textes (Transkribus). Découverte des modules spécifiques de la TEI pour certains supports (transcriptions orales, manuscrits, etc.)

Prérequis : Niveau 1 de ce cours acquis

Compétences développées :

Compétences du niveau 1 plus :

Connaitre et savoir appliquer les normes (formats, principes FAIR, ontologies et thesaurus, etc.)

Être sensibilisé aux problématiques de sauvegarde et d'archivage des ressources numériques

Descriptif : voir ci-dessus

Logiciel et/ou matériel utilisé : Oxygen XML editor

Modalités des contrôles de connaissance :

QCM sur des questions de cours, et examen sur machine (1h) mobilisant les savoir-faire acquis pendant le cours.

Référentiel de l'offre de formation

Diplôme : Master 2 Humanités numériques

Semestre : 3

Nom de l'enseignement : Vincent Bullich

Nom du ou des enseignants : Enjeux épistémologiques, politiques, juridiques, éthiques du numérique

Objectifs de l'enseignement : Présenter les principaux enjeux scientifiques et sociétaux liés au développement des « humanités numériques »

Prérequis : Aucun

Compétences développées : Perspectives critiques sur la numérisation de la recherche en SHS, connaissances juridiques quant à l'obtention, le traitement et le stockage des données, compréhension

Descriptif : Le cours se compose de deux parties complémentaires. La première consiste en une présentation des principaux enjeux scientifiques et sociétaux liés au développement des « humanités numériques ». 3 séances seront ainsi consacrées à une mise en perspectives des « humanités numériques » avec les principaux « procès » affectant la recherche en SHS à partir du moment où elle est outillée numériquement : la technisation et la médiatisation, l'industrialisation et la marchandisation, les mutations du cadre juridique et la mise en débats sociétaux des résultats de la recherche. Cette présentation sera suivie d'un temps de travaux dirigés en vu d'exposés réalisés en binôme sur des points abordés en cours et approfondis par les étudiants eux-mêmes. Un dernier temps sera consacré à une synthèse des éléments mis en avant lors des ces exposés.

Logiciel et/ou matériel utilisé : Aucun

Modalités de contrôle des connaissances et des compétences : 1 oral (50%) et 1 dossier (50%) à réaliser en binôme.

Référentiel de l'offre de formation

Diplôme : Master 2 Humanités numériques

Semestre : 3

Nom de l'enseignement : Edition numérique

Nom du ou des enseignants : Valérie Larroche, Cédric Vignault, Geoffroy Gawin

Objectifs de l'enseignement :

- Comprendre l'écosystème de l'édition numérique scientifique
- Connaître et analyser les modes d'éditorialisation
- Savoir convertir un texte au format texte en un format e-pub
- Savoir transformer un texte au format traitement de texte en un format éditorial avec les outils et méthodes du projet MÉTOPES.

Prérequis : Pas de prérequis

Compétences développées :

Une connaissance de l'édition scientifique publique et des modes d'éditorialisation des chercheurs

Un usage de la chaîne éditoriale Métopes (édition structurée en XML)

Une capacité à convertir des formats texte en format E-pub

Descriptif :

Ce cours présente les acteurs de l'édition scientifique et plus particulièrement ceux de l'édition scientifique publique. Il présente aussi les modalités d'éditorialisation des productions des chercheurs, les formats d'édition et de publication et la chaîne éditoriale utilisée chez les éditeurs scientifiques publics.

L'édition est un processus délimité dans le temps où l'auteur, l'éditeur et son équipe éditoriale travaillent à la diffusion d'un contenu travaillé ensemble qui sera daté et stabilisé. L'éditorialisation, quant à elle, prend en compte les publications qui s'enrichissent en permanence dans un processus ouvert (Chartron, 2016).

L'édition et l'éditorialisation seront abordées d'un point de vue analytique. A l'édition numérique sera ajouté l'apprentissage de savoir-faire technique.

Logiciel et/ou matériel utilisé :

- LibreOffice : <https://fr.libreoffice.org/download/telecharger-libreoffice/>
- Word
- Xml Mind (version fournie par Métopes)
- Gimp : <https://www.gimp.org/downloads/>
- Calibre : <https://calibre-ebook.com/fr/download>
- Sigil : <https://sigil-ebook.com/>
- Pandoc : <https://pandoc.org/installing.html>
- Daisy (extension de Word) : <https://daisy.org/activities/software/wordtoepub/>

Modalités de contrôle des connaissances et des compétences :

- Production d'un document au format e-pub avec la chaîne d'édition structurée en XML (Métopes) devoir à réaliser en TD à la dernière séance (50%)
- Un dossier portant sur l'éditorialisation ou l'édition numérique (50%)

Référentiel de l'offre de formation

Diplôme : Master 2 Humanités numériques

Semestre : 3

Nom de l'enseignement : Bases de données semi-structurées

Nom du ou des enseignants : Jérôme Darmont

Objectifs de l'enseignement : Approche des problématiques liées à la gestion, au stockage et à l'interrogation des données semi-structurées

Prérequis :

- Connaissance de la modélisation conceptuelle de données
- Connaissance des bases de données relationnelles et du langage SQL

Compétences développées :

- Créer des documents XML bien formés
- Créer des documents XML valides (DTD, XML Schema)
- Concevoir et exécuter des requêtes à l'aide des langages XPath et XQuery

Descriptif :

1. Introduction : Données semi-structurées
2. Langage XML
 - Éléments
 - Attributs
 - Documents XML bien formés, documents valides
 - Schéma *Document Type Definition* (DTD)
 - XML Schema
3. Langage XQuery
 - Expressions XPath
 - Requêtes FLWOR
 - Requêtes complexes : groupement et jointures

Logiciel et/ou matériel utilisé : BaseX

Modalités de contrôle des connaissances et des compétences :

- Rendu de TD – coef. 25 %
- Épreuve sur machine (2h, documents autorisés) – Coef. 75 %

Référentiel de l'offre de formation

Diplôme : Master2 Humanités numériques

Semestre : 3

Nom de l'enseignement : Modélisation et Web de données

Nom du ou des enseignants : Geoffroy Gawin

Objectifs de l'enseignement :

- Comprendre les enjeux sociétaux et scientifiques du web de données et de ses technologies.
- Savoir mobiliser les outils et les langages du web sémantique pour décrire des données et procéder à des recoupements.

Prérequis :

- Connaissances en programmation avec des langages de script (Python, SQL).
- Connaissances en base de données relationnelles.
- Idéalement, connaissances en description et en structuration de données (XML, ontologies, thesaurus...).

Compétences développées :

- Connaître les enjeux sociétaux et scientifiques des humanités numériques et de l'open Data
- Décrire des ressources numériques selon les normes, annoter et indexer des documents.
- Connaître et savoir appliquer les normes (formats, principes FAIR, ontologies et thesaurus)
- Mettre en ligne des données ouvertes dans le web de données
- Mettre en œuvre les technologies du web
- Utilisation de l'outil de gestion de code source git (à confirmer en fonction des ressources disponibles pour cet enseignement).

Descriptif :

Les cours magistraux présenteront le web de données et les technologies qui lui sont associées, les enjeux qu'il soulève. Ces cours viseront notamment à faciliter l'adoption de la part des étudiants d'une posture critique vis-à-vis des discours technicistes tenus à propos du web de données. L'accent sera mis sur la découverte des possibilités d'usage scientifique du web sémantique en travaux dirigés. Les étudiants seront amenés à programmer avec les différents langages en lien avec le web de données et réaliseront un dossier proposant un questionnaire et une démarche mobilisant les technologies du web sémantique. Ils s'interrogeront sur les limites de ces dernières. Chaque groupe présentera lors d'un exposé ses avancées dans la mise en œuvre du projet.

Logiciel et/ou matériel utilisé :

- PC avec logiciels open source (Anaconda, Turtoise Git,...)
- Si possible du côté des établissements : serveur git.

Modalités de contrôle des connaissances et des compétences :

- dossier collectif (0,5) et exposé de groupe (0,5)

Référentiel de l'offre de formation

Diplôme : Master2 Humanités numériques

Semestre : 3

Nom de l'enseignement : Data visualisation

Nom du ou des enseignants : Nicolas Guilhot

Objectifs de l'enseignement :

Acquérir les compétences de base en visualisation de données :

- Connaître les différentes formes de visualisation (graphiques, cartes, réseaux) et leur contexte d'utilisation pertinente.
- Ne pas être dépendant d'un langage ou d'une librairie :
 - Savoir préparer les données pour la réalisation des différentes formes de visualisation avec différents langages et librairies.
 - Maîtriser les bases de création et de personnalisation des différentes formes de visualisation avec différents langages et librairies.

Prérequis :

- Installation de librairies Python et R.
- Manipulation de données de base en Python et R.

Compétences développées :

- Connaître les bases de la sémiologie graphique.
- Choisir une forme de visualisation adaptée à un jeu de données.
- Préparer les données pour la visualisation avec différents langages et librairies.
- Réaliser des graphiques et des cartes thématiques avec Python et R.
- Visualiser un réseau avec le logiciel Gephi et avec R (igraph).

Descriptif :

Le cours a pour objectif de faire acquérir les compétences de base de la visualisation des données à travers la mise en application de plusieurs formes de visualisation (graphiques, cartes, réseaux). Pour chacune de ces formes, les contextes d'utilisation pertinente et les principaux types de représentation seront étudiés à partir d'exemples à critiquer et des cas pratiques seront réalisés avec les langages Python et R.

- Introduction à la visualisation des données et à la sémiologie graphique (2h)
- Présentation des langages, librairies et du concept d'environnement virtuel (1h)
- Les graphiques : sémiologie, approche critique, cas pratiques (5h40)
- Les cartes : sémiologie, approche critique, cas pratiques (6h)
- Les réseaux : sémiologie, approche critique, cas pratiques (3h)
- Évaluations (20mn et 3h)

Logiciel et/ou matériel utilisé :

- RStudio
- Jupyter Notebook ou JupyterLab

Modalités de contrôle des connaissances et des compétences :

- Analyse critique d'une visualisation (écrit, 20 minutes)
- Réalisation d'une visualisation simple (examen sur machine, 3h)

Bibliographie :

Manuels :

- Wilke, Clause O. (2019). *Fundamentals of Data Visualization: A Prime in Making Informative and Compelling Figures*. Sebastopol, CA : O'Reilly Media, Inc. <https://www.oreilly.com/library/view/fundamentals-of-data/9781492031079/>
- Lagnel, Jean-Marie (2017). **Manuel de Data Visualisation**. Paris : Eyrolles. <https://www.eyrolles.com/Informatique/Livre/manuel-de-datavisualisation-9782100738816/>
- Healy, Kieran (2018). **Data Visualization: A Practical Introduction**. Princeton : Princeton University Press. <https://press.princeton.edu/books/hardcover/9780691181615/data-visualization>

Classiques de la sémiologie graphique :

- Tufte, Edward R. (2001). **Envisioning Information** (11e éd.). Cheshire, CT : Graphics Press. 1ère édition : 1990.
- Tufte, Edward R. (2001). **The Visual Display of Quantitative Information** (2e éd.). Cheshire, CT : Graphics Press. 1ère édition : 1983.
- Bertin, Jacques (1967). **Sémiologie graphique. Les diagrammes. Les réseaux. Les cartes**. Paris : Gauthier-Villars. Réédition : Parsi, EHESS, 1999.
- Bertin, Jacques (1977). **La graphique et le traitement graphique de l'information**. Paris : Flammarion. Réédition : Bruxelles, Zones sensibles, 2017.

Quelques sites internet :

- Le site du sociologue Kieran Healy : <https://kieranhealy.org/>
- Le site de l'historien Martin Grandjean : <http://www.martingrandjean.ch/archive-reseau-visualisation-donnees-sciences-humaines/>
- The Programming Historian : <https://programminghistorian.org/>
- VisuaLoop : Infographics & Data Visualizations : <http://visualoop.com/>
- The Data Visualisation Catalogue : <https://datavizcatalogue.com/>

Référentiel de l'offre de formation

Diplôme : Master 1 Humanités numériques

Semestre : 3

Nom de l'enseignement : Machine Learning pour l'analyse des données structurées et des réseaux

Nom du ou des enseignants : Sabine LOUDCHER et Jean-Philippe MAGUE

Objectifs de l'enseignement :

L'objectif de cet enseignement est de donner un aperçu des possibilités offertes aujourd'hui par des approches de fouille de données (machine learning) puis d'inculquer les notions essentielles pour réaliser une analyse de données structurées ou des réseaux avec :

- Principaux défis de la fouille de données, enjeux et applications
- Concepts de base en apprentissage automatique supervisé
- Concepts de base de la théorie des graphes
- Evaluation des résultats des algorithmes

Prérequis :

- Statistique exploratoire
- Réduction de dimension
- Programmation en Python avec des notebooks, librairies de science des données

Compétences développées :

- Charger un jeu de données (par ex. disponible dans l'open data)
- Sélectionner la ou les méthodes adaptées au type des données (structurées ou de réseaux) et à la problématique
- Analyser des données grâce à des techniques de statistique et d'intelligence artificielle
- Savoir programmer pour analyser et visualiser des données
- Comprendre de manière critique et réflexive les données, les méthodes et les outils numériques

Descriptif :

Le cours est divisé en quatre parties : a) introduction générale à la fouille de données supervisée, b) méthodes de classification supervisée pour des données structurées c) concepts de base de la théorie des graphes, d) analyse des données de réseaux. Les cours sont ponctués par des expérimentations réalisées directement par les étudiants à l'aide de Notebooks Python.

Logiciel et/ou matériel utilisé :

Python, librairie de science des données

Modalités de contrôle des connaissances et des compétences :

- Deux examens d'une heure sur machine

Référentiel de l'offre de formation

Diplôme : Master2 Humanités numériques

Semestre : 3

Nom de l'enseignement : Machine learning pour le traitement des données textuelles

Nom du ou des enseignants : Julien Velcin

Objectifs de l'enseignement : L'objectif est de présenter les approches récentes de traitement des données textuelles, en particulier pour résoudre des tâches de recherche d'information et de classification automatique.

Prérequis :

- Programmation Python
- Bases de l'apprentissage automatique
- Bases du traitement des données textuelles (chargement des données, structuration du jeu de données, segmentation des phrases)

Connaissances à acquérir :

- Comment représenter les données textuelles
- Avantages et limites des différentes approches
- Différentes approches (classiques et récentes) pour la recherche d'information
- Modèles et algorithmes adaptés aux données textuelles pour la classification

Compétences développées :

- Encoder des textes pour des traitements ultérieurs
- Développer un moteur de recherche d'information
- Déployer des algorithmes de classification automatique

Descriptif :

Ce cours vise principalement à comprendre comment les textes peuvent être « encoder » dans une machine afin d'en capturer le sens et de résoudre des tâches comme la recherche d'information et la classification automatique. Je présenterai les approches classiques basées sur des sacs de mot, mais également les approches plus modernes basées sur des réseaux de neurones profonds (*deep learning*). La présentation des concepts sera complétée par une mise en pratique sur des cas d'application concrets.

Logiciel et/ou matériel utilisé :

Développement en Python. Environnement Anaconda en local et utilisation de solution en ligne pour les traitements plus lourds (type google colab).

Modalités de contrôle des connaissances et des compétences :

- Examen écrit 1h30
- Dossiers rendus à l'issue des TP