

# Projet Documents Structurés

Master 2 Plurital

Date limite :

## 1 Description du projet

Explorer une problématique qui vous intéresse en vous servant des différents thèmes abordés en cours. La problématique est « personnelle », elle répond à un sujet pour lequel vous avez des affinités parmi des domaines variés (sciences sociales, linguistiques, etc.). Cette exploration mène à une présentation de votre démarche et des résultats sur un site web statique.

L'entrée du projet est les données, à vous de récupérer les données nécessaires afin de répondre à votre problématique. La collecte des données est libre (attention à respecter les droits), il vous est possible d'obtenir des données déjà en XML ou bien dans un autre format. Néanmoins, quelque soit vos données, il faut les convertir en XML si ces dernières ne le sont pas déjà, ou bien les remanier pour ne garder que les informations pertinentes sur votre sujet.

Le résultat sera présenté sur un site web statique ayant pour structure au moins deux page,

- une page présentant votre projet et généré avec l'aide d'une feuille xslt
- un autre page permettant de naviguer parmi vos résultats

Bien sûr, il vous est possible d'ajouter d'autres pages.

La présentation des résultats est libre, c'est à dire que rien ne vous est imposé, à vous de choisir la visualisation la plus pertinente (une carte pour des données géographiques, des graphiques pour des données chiffrées, etc.). Plusieurs librairies sont disponibles pour vous aider comme d3.js ou bien echartsjs.

Le projet s'effectue en groupe de deux ou trois. L'objectif est d'effectuer un travail collaboratif à l'aide de *git* et de suivre l'évolution du projet avec une *kanban board*.

Ce projet sera découpé en quatre étapes, plus la gestion de projet.

1. structuration de l'environnement de travail, récupération des données et pré-traitement de celles-ci
2. analyse et sélection des données pertinentes, modélisation XML
3. établissement d'une grammaire de validation (dtd/ relaxng)
4. formatage des données et transformation xslt pour présenter votre projet sur le site web

## 2 Structuration

Cette étape de préparation comprend les points suivants :

- création manuelle de l'architecture (dossiers/fichiers) initiale ;
- récupération des données ;

- nettoyage des données ;

## 2.1 Architecture

Votre projet sera hébergé sur votre Github. L'arborescence à respecter est la suivante :

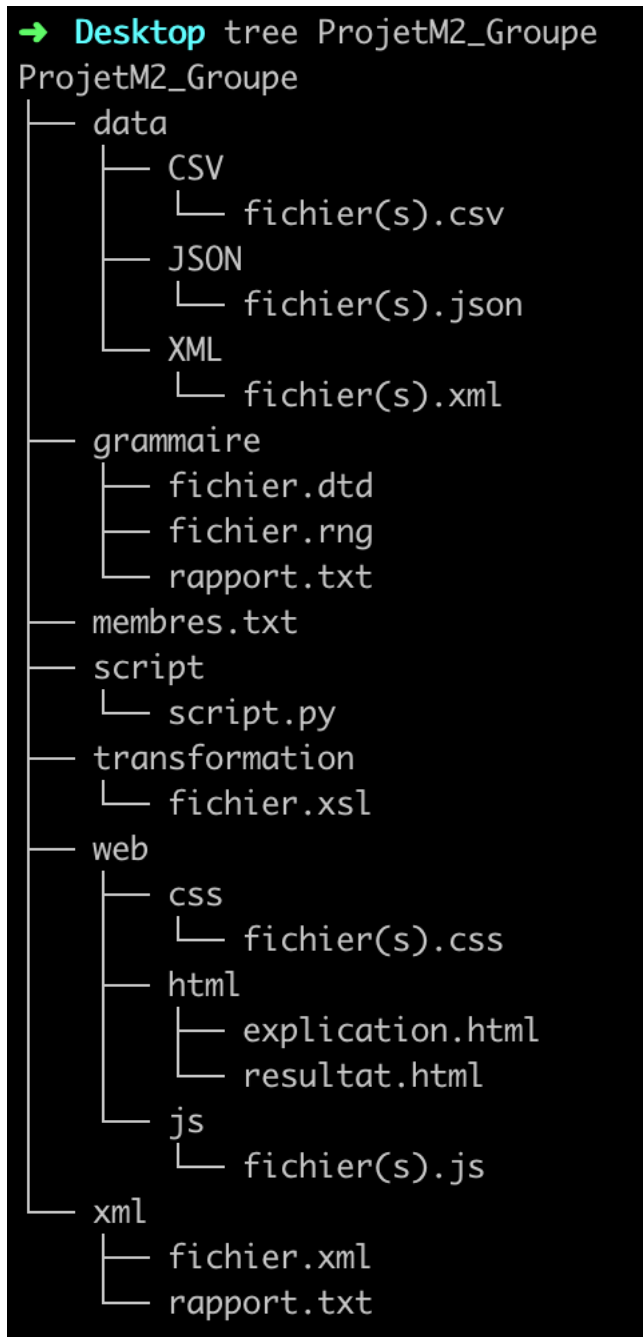


Figure 1: Architecture du projet

## 2.2 Présentation de votre projet

Dans votre fichier de « result.html », issu d'une transformation xslt, vous présenterez votre corpus (acquisition, encodage, taille, etc.), les prétraitements effectués et une explication des différents scripts utilisés pour la récupération et le traitement des données.

## 3 Modélisation XML et analyse

Une première étape correspond à la structuration de vos données en XML, dans un encodage respectant votre problématique. La pertinence de la modélisation est prise en compte.

La seconde étape est la validation de votre document xml avec une grammaire DTD et un schéma RelaxNG. Si votre document n'est pas valide et bien formé, vous serez très fortement pénalisé.

## **4    Partie web**

Fruit de votre projet, il héberge vos résultats. Le contenu principal présentant votre projet est issu d'une transformation xslt, la partie résultat est libre.