

I Descriptif du projet

Exploration thématique : Le but de ce projet est d'étudier une thématique choisie par vous, et plus particulièrement de répondre à une problématique émergente sur ce thème. Vous êtes amené à construire un site web statique qui servira à la présentation de vos données de travail. Le projet total est noté sur 50 pts.

Le type des données d'entrée est libre (csv, xml, json, html ...) l'important est d'établir une corrélation entre ces données qui vous aidera à répondre à la problématique. Pour cela on utilisera XML comme pivot pour modéliser les données.

Le site web en sortie sera présenté sous le format de deux pages principales connectées ayant les caractéristiques suivantes :

- Un fichier.html présentant votre projet et généré grâce à une feuille de style xslt (processeur saxon) mais qui contiendra le descriptif :
 - Du thème et de vos données,
 - Des méthodes utilisées pour chacune des étapes (cf infra),
 - Des résultats, des difficultés rencontrées (ce passage peut aussi être abordé dans les méthodes), une critique constructive de votre travail (améliorations possibles, avantages/inconvénients, ... etc),
- Un fichier.html contenant un tableau html permettant par le biais de liens de consulter toutes les données du projet.

Ce projet se découpe en quatre étapes :

1. Structuration de l'environnement de travail, récupération des données et pré-traitement de celles-ci ⇒ (10 pts)
 2. Analyse et sélection des données pertinentes, modélisation XML ⇒ (10 pts)
 3. Établissement d'une grammaire de validation (dtd/relaxng) ⇒ (10 pts)
 4. Formatage des données et transformation pour le tableau web après création de celui-ci ⇒ (20 pts)
- Chacune des étapes va être décrite ci-dessous.

II Structuration 10 pts

Cette étape de préparation comprend les points suivants :

- Création manuelle de l'architecture (dossiers/fichiers) initiale,
- Récupération des données,
- Nettoyage des données

II.1 Architecture initiale

Initialement vous créerez la structure de dossiers/fichiers correspondant à l'exemple suivant :

Le dossier principal contiendra un README explicatif de tous les éléments contenus (avec les sources si vous n'êtes pas producteur de la ressource)

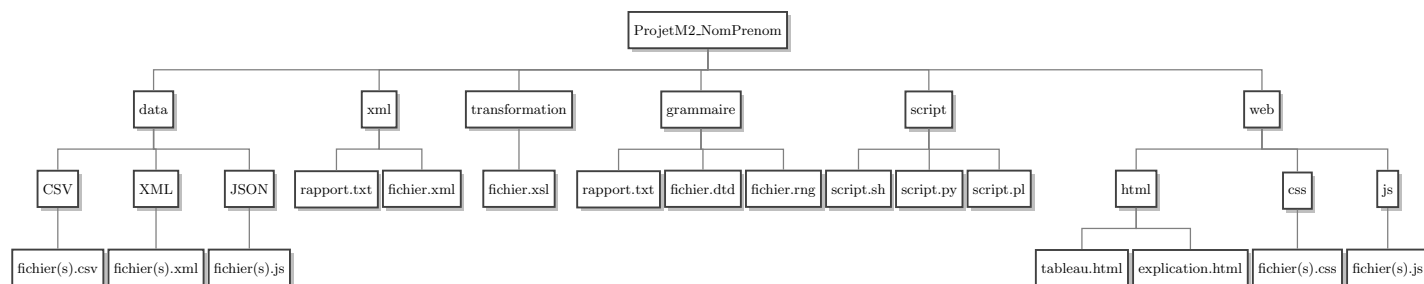


FIGURE 1 – Architecture dossier(s)/fichier(s) du projet

II.2 Récupération des données

Vous détaillerez cette étape dans votre page web explicative. ⇒ Détail du corpus avec par exemple :

- taille initiale (Mo/Go ? Nombre de mots ? nombre de balises (xml) ? colonnes/lignes (csv) ?),
- taille finale après traitement (idem),
- % de perte si il y a lieu lors du traitement,
- échelle de temps,
- type de données.

II.3 Nettoyage des données

Cette partie sera, elle aussi, abordée dans votre descriptif. Si vos données sont bruitées, vous expliquerez quelle a été votre méthode de traitement et quelle en est la perte résultante.

II.4 Détail des scripts

La définition de l'architecture de votre projet lui permet de fonctionner de la même manière quelque soit l'ordinateur où on l'utilisera. Il convient donc de faire appel aux chemins de manière relative dans vos scripts. Il sera clairement explicité dans les synopsis des scripts quels sont les modules/bibliothèques à installer et les conditions d'utilisation.

III Analyse et modélisation XML 10 pts

Dans cette étape vous modéliserez vos données selon un modèle clairement défini en xml. Ce modèle doit vous permettre de répondre à la problématique. En outre votre document doit être valide selon les standards xml.

IV Grammaire 10 pts

Établissez les grammaires (dtd et relax-ng) vous permettant de contraindre votre document. Le dossier grammaire contiendra le résultat des tests de validités dans un rapport (redirection vers fichier txt par exemple).

V Web 20 pts

Cette partie comprend deux étapes, la transformation de vos données puis l'utilisation de celles-ci dans votre site web.

Comme mentionné (cf supra) votre site web se composera à minima de deux pages web, la première étant un récapitulatif de votre projet (résultat d'une transformation xslt).

La seconde se présentera sous forme de tableau (avec en-tête explicative) et permettra d'accéder à toutes vos données de travail y compris les modélisations que vous aurez choisi pour présenter vos résultats.

Attention :

- Le choix de représentation doit illustrer votre approche pour répondre à la problématique que vous vous êtes posée.
- L'utilisation d'éléments techniques vus en cours (entités, etc) est un plus dans la notation.