Données semi-structurées

A. U: 2022/2023

TP N° 01 (Partie 01)

Objectifs

L'objectif de cette série de TP est de :

- Faire un rappel sur le langage HTML : le langage HTML est un autre langage basé sur les balises, tout comme XML. Ce langage a fait l'objet de plusieurs travaux durant les deux premières années de formation. Dans cette série, vous allez revoir la notion de document HTML en créant un document riche en éléments.
- Faire un rappel sur la notion de DOM : Document Object Model est le modèle principal pour la manipulation des éléments d'une page (ou document HTML). Dans cette série, l'objectif est d'analyser un document HTML et d'extraire sa structure hiérarchique.

Travail à réaliser

Créer un document HTML

La premier étape dans cette série consiste à créer un document HTML. Pour pouvoir l'exploiter plus tard, le document à créer doit être riche en éléments :

- Le document doit présenter une structure claire : la streuture d'une page HTML se compose généralement des composants suivants :
 - o Entête,
 - Barre de navigation,
 - Panneau de navigation (droite, gauche, ou bien les deux à la fois),
 - o Contenu principal : ce contenu peut être aussi structuré en entête, corps et pieds.
 - Pieds de la page.
 - Cette structuration peut utiliser des tableaux (appoche ancienne) ou bien des divs stylées (approche plus moderne)
- Le document doit contenir une table et afficher une image.
- Vous êtes libres à ajouter les éléments de styling (CSS) que vous jugez nécessaires.

Analyse du document HTML

Un dcoument HTML se compose d'un ensemble de noeuds. Ces noeuds sont structurés sous forme hiérarchique. Cette structure repose sur le modèle DOM. Pour pouvoir manipuler dynamiquement le contenu d'une page HTML, cette structure hiérarchique peut être utilisée.

Dans cette série, l'objectif est d'analyser le document créé dans la première étape et d'extraire sa structure hiérarchique.