

TD2 : introduction à XML et XPath

Télécharger l'archive *TD2.zip* depuis l'espace-cours Coursus et décompresser cette archive dans un répertoire personnel. Le fichier de données au format XML s'appelle *films.xml*, et le fichier de code Python à modifier est *td2.py*.

Le fichier de données *films.xml* contient des données sur des films et des artistes (acteurs et metteurs en scène) avec la structure d'éléments suivante :

- <FILMS> : élément racine du document
- <FILM> : élément représentant un film
 - <TITRE> : élément « titre » du film
 - <ANNEE> : élément « années de sortie » du film
 - <GENRE> : élément « genre » du film
 - <MES> : élément « metteur en scène du film. Cet élément contient un attribut *idref* qui fait référence à un attribut *id* d'un élément <ARTISTE>
 - <ROLES> : élément contenant la liste des élément <ROLE> du film
 - <ROLE> : élément de « rôle » du film. Cet élément contient un attribut *idref* qui fait référence à un attribut *id* d'un élément <ARTISTE>
 - <RESUME> : élément « résumé » du film
- <ARTISTE> : élément représentant un artiste. Cet élément contient un attribut *id* identifiant de façon unique l'artiste.
 - <ACTNOM> : élément « nom » d'un artiste
 - <ACTPNOM> : élément « prénom » d'un artiste
 - <ANNEENAISS> : élément « année de naissance » d'un artiste

Préambule :

- Écrire le code permettant de lire le fichier *films.xml* et de récupérer l'élément racine de l'arbre XML correspondant. Afficher cet arbre dans la console.
- Écrire le code Python permettant d'obtenir la liste des éléments <FILM> de l'arbre XML

Exercice 1 :

1. Écrire une fonction *afficheTexte()* permettant d'afficher le **contenu textuel** d'une liste d'éléments XML passée en paramètre, et de tous leurs sous-éléments. Tester sur la liste des élément <FILM> obtenue en préambule.

Exemple d'affichage :

Titanic

1997

Drame

None

None

Rose DeWitt Bukater

Jack Dawson

Conduite par Brock Lovett, une expédition américaine fouillant l'épave du Titanic remonte à la

surface le croquis d'une femme nue. Alertée par les médias la dame en question, Rose DeWitt

Bukater, aujourd'hui centenaire, rejoint les lieux du naufrage, d'où elle entreprend de conter le

récit de son fascinant, étrange et tragique voyage...

Sacrifice

1986

Drame

None

None

None

Dans l'affichage obtenu, à quoi correspondent les lignes vides ? De même, à quoi correspondent les valeurs *None* ?

2. Écrire une fonction *afficheResultatRequete()* permettant d'afficher le code (XML ou texte) d'une liste d'éléments résultant d'une requête Xpath (cette liste sera passée en paramètre de la fonction). Indication : on pourra tester si les éléments dans cette liste sont des éléments XML en utilisant la fonction [*isinstance\(\)*](#) de Python ; la classe d'un élément XML dans *etree* est *etree.Element*.

Tester sur la liste des élément <FILM> obtenue en préambule

3. Écrire une fonction permettant d'afficher les informations d'une liste d'éléments <FILM>

passée en paramètre. Chaque élément <FILM> sera affiché de la façon suivante (attention, il faudra gérer le cas où le résumé est absent) :

Titre : Le crime était presque parfait

Année : 1954

Genre : Policier

Rôles : Margot Wendice, Tony Wendice

Résumé : En apprenant que sa femme Margot a un amant, Tony décide de la tuer. Afin de ne pas être accusé du meurtre et d'hériter de sa fortune, il oblige Lesgate, un ancien ami, à commettre le crime. Mais lors de l'agression, Margot parvient à se défendre et tue Lesgate. Habilement, Tony persuade la police que ce n'était pas de la légitime défense.

Exercice 2 :

Écrire le code Python permettant d'afficher les informations suivantes **en utilisant une seule requête Xpath** (on pourra utiliser la fonction *afficheResultatRequete()* pour visualiser les résultats lorsque cela est approprié) :

1. La liste des noms de tous les artistes ;
2. Même question mais avec affichage des balises ;
3. Afficher le titre des films sortis après 1990 ;
4. Afficher le résumé du film Blade Runner ;
5. Afficher le dernier film du document ;
6. Afficher le dernier film du document paru en 1990 ; *Indication : on peut cascader des prédicats pour appliquer plusieurs filtres successivement*
7. Afficher les titres des films qui ont un résumé, puis ceux qui n'en ont pas ;
8. Afficher les titres des films vieux de plus de trente ans ;
9. Afficher les titres des films qui contiennent un « V » (utiliser la fonction contains()) ;
10. Afficher les titres des films dont le texte du résumé contient le mot "amour" ;
11. Afficher le nombre de films ;
12. Afficher le nombre de drames ;
13. Donner les noms des acteurs qui ont joué dans Vertigo ; *Indications :*
 - *on peut utiliser une expression Xpath à l'intérieur d'un prédicat*
 - *on peut comparer la valeur d'un élément à une liste de valeurs*
14. Afficher le nom du metteur en scène de Vertigo ;
15. Afficher tout les films qui ont le même metteur en scène que Vertigo ;
16. Afficher les titres des films où Bruce Willis a joué (comme acteur) ;
17. Afficher le rôle de Harvey Keitel dans le film Reservoir dogs.