Données Semi-Structurées:

DOM et requête XPath avec Python

Enseignant: *Dario COLAZZO*
Chargée de TD/TP: *Beatrice NAPOLITANO*

01 Février 2021

1 Dom et Python

DOM permet de manipuler un document XML en mémoire comme un arbre d'objets représentant les noeuds du document. Les interfaces standard DOM sont les suivantes.

Node	Interface de base des noeuds
NodeList	Séquence de noeuds
Document	Représente un document complet
Element	Elément de la hiérarchie
Attr	Noeud représentant un attribut d'un noeud
Comment	Noeud repréntant un commentaire
Text	Noeud de données textuelles

Le module *minidom* est une mise en oeuvre de DOM pour Python. Il fournit toutes les interfaces de base DOM et un parser de fichier (ou de chaîne) XML. Il est disponible en standard dans l'environnement.

Documentation : https://docs.python.org/2/library/xml.dom.html

2 Example

Considèrez le document XML suivante (*carnet.xml*) dècrivant un carnet d'adresses :

```
1: <?xml version="1.0" encoding="UTF-8"?>
2: <carnet>
3:  <address name="Beatrice Napolitano" id="_1">
4:  <company>Paris-Dauphine</company>
5:  <phone>06 12345678</phone>
6:  </address>
7:  <address id="_2">
8:  <company>Paris-Dauphine</company>
9:  <phone>07 12141618</phone>
10:  </address>
11: </carnet>
```

Considérez la fonction Python suivante qui imprime les identifiants des éléments qui ont un nom :

```
from xml.dom.minidom import parse

def getId():
        dom = parse("carnet.xml")
        print(dom.hasChildNodes())
        for n in dom.getElementsByTagName("address"):
                if (n.hasAttribute("name")):
                        print(n.getAttribute("id"))
```

3 Exercices

Exercice 1 *Testez la fonction Python de l'Exemple 2 :*
- *création de document carnet.xml ;*
- *lecture de document et test de fonction getId() ;*
- *implémentation d'une fonction imprimant tout les numéros de téléphone présents dans le carnet.*

Exercice 2 *Écrivez un document XML Films.xml basé sur la DTD de l'exercice 8 du TP1, et des fonctions Python pour répondre aux requêtes du même exercice (de 1 à 9).*

```
1: <!DOCTYPE FILMS [
2: <!ELEMENT FILMS (FILM+, ARTISTE+)>
3: <!ELEMENT FILM (TITRE, GENRE, PAYS, MES, ROLES, RESUME?)>
4: <!ELEMENT TITRE (#PCDATA)>
5: <!ATTLIST FILM Annee CDATA #REQUIRED>
6: <!ELEMENT GENRE (#PCDATA)>
7: <!ELEMENT PAYS (#PCDATA)>
8: <!ELEMENT MES EMPTY>
9: <!ATTLIST MES id_mes IDREF #IMPLIED>
10: <!ELEMENT ROLES (ROLE*)>
11: <!ELEMENT ROLE (PRENOM, NOM, INTITULE)>
12: <!ELEMENT PRENOM (#PCDATA)>
13: <!ELEMENT NOM (#PCDATA)>
14: <!ELEMENT INTITULE (#PCDATA)>
15: <!ELEMENT RESUME (#PCDATA)>
16: <!ELEMENT ARTISTE (ACTNOM, ACTPNOM, ANNEENAISS)>
17: <!ATTLIST ARTISTE id_art ID #REQUIRED>
18: <!ELEMENT ACTNOM (#PCDATA)>
19: <!ELEMENT ACTPNOM (#PCDATA)>
20: <!ELEMENT ANNEENAISS (#PCDATA)>
21: ]>
```

1. *La liste des titres de films.*
2. *Les titres des films parus en 1990*
3. *Le résumé d'Alien*
4. *Titre des films avec Bruce Willis*
5. *Quels films ont un résumé ?*
6. *Quels films n'ont pas de résumé ?*
7. *Donner les titres des films vieux de plus de trente ans.*
8. *Quel rôle joue Harvey Keitel dans Reservoir dogs ?*
9. *Quel est le dernier film du document ?*