

Séance 2

Interlude... où en sommes-nous ?

Données ouvertes : des données brutes disponibles

La principale motivation de l'open data est la transparence des données accessibles à tout le monde. L'une des motivations est d'également permettre aux développeurs d'utiliser ses données pour les intégrer dans des applications et agrandir la visibilité des open-data

Données ouvertes, des histoires de formats...

- En naviguant sur les différentes API, plusieurs formats sont proposés régulièrement (json, xml, csv), mais il existe d'autres formats moins répandus pour ce type de ressource : php, xmlfm & jsonfm (version HTML de XML ou JSON), KML (format XML pour du contenu géographique)
- Les données XML sont étiquetées dans des balises
- Les données JSON sont étiquetées sous forme de booléens, entiers, chaîne de caractères compris dans des objets ou de tableaux
- Les données CSV sont étiquetées sous forme de texte séparé par des virgules, qui forment un tableau traité par un logiciel tel que Excel
- Exemple de données en XML avec l'API OMBD :
 - **URI** : www.omdbapi.com/?apikey=7403b6e6&r=XML&t=Fight%20Club
 - **Paramètres** : r = format de données, t = titre exact du film
 - **Autres paramètres possible** : i = id IMBD du film, s = chaîne de caractères comprise dans le titre, y = année de sortie, ...
 - **Réponse** :

```
<root response="True">
<movie title="Fight Club" year="1999" rated="R" released="15 Oct 1999" runtime="139 min"
genre="Drama" director="David Fincher" writer="Chuck Palahniuk, Jim Uhls" actors="Brad Pitt,
Edward Norton, Meat Loaf" plot="An insomniac office worker and a devil-may-care soap maker
form an underground fight club that evolves into much more." language="English"
country="Germany, United States" awards="Nominated for 1 Oscar. 11 wins & 38 nominations
total"
poster="https://m.media-amazon.com/images/M/MV5BMmEzNTkxYjQtZTc0MC00YTVjLTg5ZT
EtZWwOWVlYzY0NWlWxkEyXkFqcGdeQXVyNzkwMjQ5NzM@._V1_SX300.jpg"
metascore="66" imdbRating="8.8" imdbVotes="1,958,640" imdbID="tt0137523" type="movie" />
</root>
```

Et l'interopérabilité dans tout ça ?

Les données récupérées peuvent avoir plusieurs utilisations, comme pour les données de suivi de l'épidémie Covid-19, trié selon l'hôpital ou le département. Les données proposés par l'API OMBD peuvent servir à obtenir des informations sur nos films préférés.

Séance 2 - Un type de données structurées XML

Jouons un peu pour découvrir XML

- Un fichier DTD (**D**ocument **T**ype **D**efinition), est une description du document XML (nom, type, répétition, ordre, ...)
- Cette dtd contient 9 éléments au total
- L'élément CD contient 5 éléments minimum (Titre, Interprète, Genre, Piste+, Pochette)
- Correction fichier dtd :
-

```
<!ELEMENT CDs (CD)>
<!ELEMENT CD (Titre, Interprete, Genre, Piste+, Pochette)>
<!ELEMENT Piste (Titre, Compositeur+, Duree)>
<!ELEMENT Titre (#PCDATA)>
<!ELEMENT Interprete (#PCDATA)>
<!ELEMENT Genre (#PCDATA)>
<!ELEMENT Compositeur (#PCDATA)>
<!ELEMENT Duree (#PCDATA)>
<!ELEMENT Pochette (Urlface, Urldos?)>
<!ELEMENT Urlface (#PCDATA)>
<!ELEMENT Urldos (#PCDATA)>
<!ELEMENT site (#PCDATA)>
```

- Le fichier XML n'est pas structuré correctement, voici les différentes corrections :
 - la balise racine **CDS** est manquante -> ajout de la balise racine **CDS**
 - la balise **CD** est manquante pour les 3 CD-> ajout de la balise **CD**
 - Dans les CD 1 & 2, il manque l'élément **site**
 - Les attributs **urlface & urldos** sont remplacés par des sous-éléments de Pochette
 - Dans le 1^{er} CD, au lieu de 3 interprètes, il ne faut qu'1 interprète
 - Dans le 2^{ème} CD, 1^{ère} & 2^{ème} piste, il manque la durée -> ajout de l'élément **durée**
 - Dans le 3^{ème} CD, Interprète & Genre sont inversés
 - Dans le 3^{ème} CD, 1^{ère} & 2^{ème} piste, il manque la durée -> ajout de l'élément **durée**

Lien Github : https://github.com/sebastienklaus/Interoperabilite/tree/master/seance_2