

# TP1 - M2 GIL - BD NoSQL pour le Big Data XPath

25 septembre 2016  
Lina Soualmia

L'objectif de ce TP est de manipuler des documents XML à l'aide de Bases XML Natives que vous devez installer sur vos machines. Vous pouvez également utiliser baseX.

## 1 Préambule

Vous pouvez au choix installer le logiciel eXist (Système open source de gestion native de bases de données XML) que vous pouvez télécharger sur <http://exist-db.org/exist/apps/homepage/index.ht> ou encore BaseX <http://www.inf.uni-konstanz.de/dbis/basex/> (prendre le .jar).

### 1.1 Import

L'import permet de créer des '*collections*'. Une collection peut être utilisée pour stocker un ensemble de documents partageant le même schéma. Dans le menu '*Browse collections*' du menu admin vous pouvez y accéder. Vous pouvez également créer une nouvelle collection.

1. créez une collection **movies**
2. ajoutez à cette collection le document **movies.xml** (disponible dans les 'Documents partagés' de votre liste de diffusion).

### 1.2 Expressions XPath, XQuery via Sandbox

1. Testez l'expression suivante :  
`/movies`  
Elle permet d'afficher le contenu des éléments de type **movies** existant dans toutes les collections stockées sous l'élément racine d'eXist `/db`.
2. Vous pouvez restreindre la recherche à des documents spécifiques (ou un ensemble de documents) avec la fonction `document()`.  
Testez : `document('/db/movies/movies.xml')/movies/movie[year=2005]`
3. Pour restreindre la recherche à une collection, il faut utiliser la fonction `collection()` :  
`collection('movies')/movies/movie[year=2005]`

## 2 Exercices

Il vous est demandé un compte-rendu (personnel) de ce TP et à envoyer avant la prochaine séance avec comme objet TP1 NOM à : [bd.m2.nosql@gmail.com](mailto:bd.m2.nosql@gmail.com).

## 2.1 XPath

Récupérez les documents `movies.xml` et `movies_refs.xml` ('Documents partagés' de votre liste de diffusion) et insérez-les dans votre SBDG XML. Dans le premier document, chaque film (`movie`) est représenté comme un seul élément, incluant les noms des acteurs (`actor`) et des réalisateurs (`director`). Dans le second document, chaque movie est composé de deux listes : une liste de films et une listes d'acteurs, avec des références. Pour chaque document, exprimez les requêtes suivantes en XPath :

1. Tous les éléments titre (`title`).
2. Tous les titres des films (la valeur texte des éléments `title`).
3. Titres des films parus après 2000.
4. Le résumé (`summary`) du film 'Spider-Man'.
5. Qui est le réalisateur (`director`) de Heat ?
6. Titres des films avec Kirsten Dunst.
7. Quels films ont un résumé ? (`summary`)
8. Quels films n'ont pas de résumé ?
9. Titres des films parus il y a plus de 5 ans ?
10. Quel est le rôle de Clint Eastwood dans Unforgiven ?
11. Quel est le dernier (`last`) film du document ?
12. Titre du film qui précède immédiatement Marie Antoinette dans le document ?
13. Récupérez les films dont le titre contient 'V' (fonction `contains ()`).
14. Récupérez les films qui ont exactement 3 acteurs (fonction `count()`).