

TP N° 3: XML (XPath)

Exercice 1 : Proposer des requêtes XPath permettant de constituer les ensembles suivants à partir du fichier 'article.xml' :

```
<?xml version="1.0" ?>
<article>
  <titre> A propos des DTD </titre>
  <corps >
    <p> Les DTD, <important>c'est facile</important>, il suffit de lire le cours
    de Marc Tommasi <cite ref="siteMarc" />.</p>
    <p> Il y a aussi de bons livres <cite ref="AmannRigaux"/>. </p>
  </corps>
  <references>
    <reference code="siteMarc">
      <nom>Site de Marc Tommasi</nom>
      <url> http://www.grappa.univ-lille3.fr/~tommasi/Homepage/SurDtd.html</url>
    </reference>
    <reference code="AmannRigaux">
      <intitule>Comprendre XSLT</intitule>
      <auteur>Bernd Amann</auteur>
      <auteur>Philippe Rigaux</auteur>
      <date>2002</date>
      <editor>O'Reilly</editor>
    </reference>
  </references>
</article>
```

1. Le nœud racine ;
2. Les éléments *reference*;
3. Tous les descendants de la racine ;
4. Tous les attributs ;
5. Tous les nœuds de type texte ;
6. Les contenus de tous les paragraphes ;
7. Les paragraphes contenant des aspects importants ;
8. La référence portant le code *sitemarc* ;
9. Tous les deuxièmes auteurs ;
10. Le nombre d'auteur décrit dans le document XML ;
11. Les deux premiers éléments fils de chaque référence ;
12. Le dernier élément fils de chaque référence.
13. Tous les références dont l'éditeur est O'Reilly.

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Relizane

Département Informatique
Données Semi Structurées : L3 - SI

Année académique 2021/2022
Responsable du module : S. Bella

14. Tous les références ayant un URL et non pas un auteur.

Exercice 2 : Donner la description de chaque requête XPath suivante à partir du fichier 'actu.xml' :

```
<?xml version="1.0" ?>
<actu>
  <breve theme="actu" langue="fr" date="7 décembre 2004">
    <titre> Les députés renoncent au CV anonyme </titre>
    <texte>
      Les députés ont achevé lundi, après plus de deux
      semaines de débat marathon, l'examen du projet de loi
      de cohésion sociale. Ils ont abandonné le CV anonyme,
      mesure à laquelle le gouvernement et l'UMP étaient
      hostiles en l'état.
    </texte>
    <url href="http://www.reuters.fr/locales/c_newsArticle
      .jsp?type=topNews&localeKey=fr_FR&storyID=7007458">
      Reuters </url>
    <url href="http://news.tf1.fr/news/economie/0,,3189581
      ,00.html">TF1 </url>
  </breve>
  <breve theme="sport" langue="fr" date="7 décembre 2004">
    <titre> Décès de Raymond Goethals </titre>
    <texte>
      L'entraîneur belge Raymond Goethals, décédé lundi à
      Bruxelles, suscitait la sympathie grâce à une
      truculence et une gouaille incomparables qui en
      faisaient un bon client des médias.
    </texte>
    <photo src="http://robert93.free.fr/staff/ goethals01.
      jpg" />
    <url href="http://www.reuters.fr/locales/c_newsArticle
      .jsp?type=sportsNews&localeKey=fr_FR&storyID=7007386">
      Reuters </url>
  </breve>
  <breve theme="actu" langue="fr" date="7 février 2006">
    <titre>Mobilisation contre le CPE </titre>
    <texte>
      Manifestations en ce moment dans toutes les grandes
      villes de France.
    </texte>
  </breve>
</actu>
```

1. `//*` ou `/descendant::*`
2. `//@href`
3. `//breve/titre/text()` ou `/descendant::breve/child::titre/child::text()`
4. `//breve[@theme="sport"]`
5. `//breve[photo]`
6. `//breve[starts-with(titre/text(),"D")]`
7. `//breve[contains(texte/text(),"marathon")]`
8. `//breve[@theme="actu"]/following::*`
9. `//breve[2]/titre`
10. `//breve[@date="7 février 2006"]/texte/text()`
11. `//breve[count(./url)=0]`
12. `//*[@theme][2]/@theme`
13. `/descendant::titre[position()=1]`
14. `count(//breve[1]/url)`

Solution TP N° 3: XML (XPath)

Exercice 1 : Proposer des requêtes XPath permettant de constituer les ensembles suivants à partir du fichier 'article.xml':

```
<?xml version="1.0" ?>
<article>
  <titre> A propos des DTD </titre>
  <corps >
    <p> Les DTD, <important>c'est facile</important>, il suffit de lire le cours
    de Marc Tommasi <cite ref="siteMarc" />.</p>
    <p> Il y a aussi de bons livres <cite ref="AmannRigaux"/>. </p>
  </corps>
  <references>
    <reference code="siteMarc">
      <nom>Site de Marc Tommasi</nom>
      <url> http://www.grappa.univ-lille3.fr/~tommasi/Homepage/SurDtd.html</url>
    </reference>
    <reference code="AmannRigaux">
      <intitule>Comprendre XSLT</intitule>
      <auteur>Bernd Amann</auteur>
      <auteur>Philippe Rigaux</auteur>
      <date>2002</date>
      <editor>O'Reilly</editor>
    </reference>
  </references>
</article>
```

1. Le nœud racine ;
2. Les éléments *reference*;
3. Tous les descendants de la racine ;
4. Tous les attributs ;
5. Tous les nœuds de type texte ;
6. Les contenus de tous les paragraphes ;
7. Les paragraphes contenant des aspects importants ;
8. La référence portant le code *sitemarc* ;
9. Tous les deuxièmes auteurs.
10. Le nombre d'auteur décrit dans le document XML ;
11. Les deux premiers éléments fils de chaque référence ;
12. Le dernier élément fils de chaque référence.
13. Tous les références dont l'éditeur est O'Reilly.
14. Tous les références ayant un URL et non pas un auteur.

Solution :

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Relizane

Département Informatique
Données Semi Structurées : L3 - SI

Année académique 2021/2022
Responsable du module : **S. Bella**

1. /* ou /child::*
2. //references/reference ou /descendant::reference ou //reference
3. //*//node() ou /child::*//descendant::node()
4. //@* ou /descendant::*//attribute::*
5. //text() ou /descendant::text()
6. //p/text() ou /descendant::p/child::text()
7. //p[important] ou /descendant::p[child::important]
8. //reference[@code='siteMarc'] ou /descendant::reference[attribute::code='siteMarc']
9. //*//auteur[position()=2] ou /descendant::*//child::auteur[position()=2]
10. count(/child::reference/child::auteur) ou count(/reference/auteur) ou count(/descendant::auteur) ou count(/auteur)
11. descendant::reference/*[position()=1 or position()=2] ou //reference/*[position()<=2]
12. descendant::reference/*[position()=last()] ou //reference/*[last()]
13. /article/references/reference[./editor="O'Reilly"]
14. //reference[url]
15. //reference[url and not(auteur)]

Exercice 2 : Donner la description de chaque requête XPath suivante à partir du fichier 'actu.xml' :

```
<?xml version="1.0" ?>
<actu>
  <breve theme="actu" langue="fr" date="7 décembre 2004">
    <titre> Les députés renoncent au CV anonyme </titre>
    <texte>
      Les députés ont achevé lundi, après plus de deux semaines de débat
      marathon, l'examen du projet de loi de cohésion sociale. Ils ont abandonné
      le CV anonyme, mesure à laquelle le gouvernement et l'UMP étaient hostiles
      en l'état.
    </texte>
    <url
      href="http://www.reuters.fr/locales/c_newsArticle.jsp?type=topNews&localeKey
        =fr_FR&storyID=7007458"> Reuters
    </url>
    <url href="http://news.tf1.fr/news/economie/0,,3189581,00.html">TF1
    </url>
  </breve>
  <breve theme="sport" langue="fr" date="7 décembre 2004">
    <titre> Décès de Raymond Goethals
    </titre>
    <texte>
      L'entraîneur belge Raymond Goethals, décédé lundi à Bruxelles, suscitait la
      sympathie grâce à une truculence et une gouaille incomparables qui en
      faisaient un bon client des médias.
    </texte>
    <photo src="http://robert93.free.fr/staff/goethals01.jpg" />
    <url href="http://www.reuters.fr/locales/c_newsArticle.jsp?type=sportsNews
      &localeKey=fr_FR&storyID=7007386"> Reuters
  </breve>
</actu>
```

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Relizane

Département Informatique
Données Semi Structurées : L3 - SI

Année académique 2021/2022
Responsable du module : **S. Bella**

```
</url>
</breve>
<breve theme="actu" langue="fr" date="7 février 2006">
  <titre>Mobilisation contre le CPE
</titre>
  <texte>
    Manifestations en ce moment dans toutes les grandes villes de France.
  </texte>
</breve>
</actu>
```

1. `//* ou /descendant::*`
2. `//@href`
3. `//breve/titre/text() ou /descendant::breve/child::titre/child::text()`
4. `//breve[@theme="sport"]`
5. `//breve[photo]`
6. `//breve[starts-with(titre/text(),"D")]`
7. `//breve[contains(texte/text(),"marathon")]`
8. `//breve[@theme="actu"]/following::*`
9. `//breve[2]/titre`
10. `//breve[@date="7 février 2006"]/texte/text()`
11. `//breve[count(./url)=0]`
12. `//*[@theme][2]/@theme`
13. `/descendant::titre[position()=1]`
14. `count(//breve[1]/url)`

Solution :

1. Sélectionner tous les éléments ;
2. Sélectionner les adresses web contenues dans le document (soit uniquement les attributs href) ;
3. Sélectionner tous les titres (de type texte) de brèves ;
4. Sélectionner les brèves sur le sport ;
5. Sélectionner les brèves qui contiennent une photo ;
6. Sélectionner les brèves dont le titre commence par la lettre D ;
7. Sélectionner les brèves dont le texte contient le mot marathon ;
8. Sélectionner tous les éléments qui suivent une brève de type actu.
9. Sélectionner le titre de deuxième brève;
10. Sélectionner le contenu du texte du brève daté en 7 février 2006;
11. Sélectionner le brève sans url ;
12. Sélectionner la valeur du deuxième thème apparaissant dans le document ;
13. Sélectionner le premier titre du document ;
14. Sélectionner le nombre d'URL dans le premier brève.