

Master 2 Humanités numériques – Bases de données semi-structurées Examen sur machine – 04/03/2024

J. Darmont – <https://eric.univ-lyon2.fr/jdarmont/>

Durée : 2 heures – Documents autorisés – Barème fourni à titre indicatif

Rendu : Fichiers DTD (.dtd) et XQuery (.xq) sur Moodle.
<https://moodle.univ-lyon3.fr/course/view.php?id=12832>

Exercice 1 : Schéma XML (10 points)

Définir une DTD permettant de valider le document publications.xml annexé dans Moodle.

Exercice 2 : XQuery (10 points)

Soient les trois documents articles.xml (articles de presse), images.xml (photos et métadonnées atenantes) et inventaire.xml (liste de bâtiments) liés au patrimoine industriel de la région lilloise (projet de recherche TECTONIQ¹). La structure de ces documents XML est donnée ci-dessous.

- + Cardinalité 1 à plusieurs
- * Cardinalité 0 à plusieurs
- ? Cardinalité 0 ou 1

articles.xml	images.xml	inventaire.xml
CORPUS	liste	inventaire
DOCUMENT+	enregistrement+	notice+
@id	@id-img	@reference
DESCRIPTION	description	edifice
AUTEUR+	titre	localisation
DATE	mot-cle*	historique
TITRE	mot-cle-geo*	description
SOUS-TITRE?	annee	etages
LANGUE	origine	dateVersement
NOTE	fichier	dateMiseAJour
TEXTE		
REFERENCES		
REF-IMG*		
REF-BAT*		

¹ <http://tectoniq.meshs.fr/>

Formuler à l'aide du langage XPath les requêtes suivantes.

1. Moyenne des notes des articles.
2. Adresses distinctes des bâtiments de l'inventaire.
3. Description complète des articles qui contiennent un sous-titre.
4. Identifiants des photos prises avant 1927 et dont la description contient le mot-clé « usine ».

Formuler à l'aide de la syntaxe FLWOR les requêtes suivantes. Privilégier la lisibilité du code et les résultats formatés en XML.

5. Titres et sous-titres des documents en langue Français-Chti.

6. Information concernant les images au format suivant :

```
<image id="I" statut="S" />
```

où **I** est l'@id-img de la photo et **S** est égal à « avant 1930 » si la photo a été prise avant 1930, « depuis 1930 » sinon.

7. Pour chaque image, donner son titre (dans un élément image) et le titre de l'article dans lequel elle apparaît (dans un élément article).

8. Titres, auteurs et dates des articles, ainsi que les fichiers images et les descriptions des édifices correspondants pour la ville de Roubaix.

9. Nombre moyen d'étages par localisation.

10. Nombre total de références (d'images et de bâtiments) par article (indiquer le titre).

Correction Exercice 1

```
<!ELEMENT publications (conference | book-chapter | other)*>

  <!ATTLIST name CDATA #REQUIRED>

  <!ELEMENT conference (author+, title, booktitle, location, month?,
    year, series?, publisher?, pages)>
  <!ELEMENT book-chapter (author+, title, booktitle, publisher, year,
    pages, note?)>
  <!ELEMENT other (author+, title, booktitle, series?, year, pages?)>

  <!ELEMENT author (#PCDATA)>
  <!ELEMENT title (#PCDATA)>
  <!ELEMENT booktitle (#PCDATA)>
  <!ELEMENT location (#PCDATA)>
  <!ELEMENT month (#PCDATA)>
  <!ELEMENT year (#PCDATA)>
  <!ELEMENT series (#PCDATA)>
  <!ELEMENT publisher (#PCDATA)>
  <!ELEMENT pages (#PCDATA)>
  <!ELEMENT note (#PCDATA)>
```

Correction Exercice 2

```
(: 1 :)
avg(//NOTE)

(: 2 :)
distinct-values(//localisation)

(: 3 :)
//DOCUMENT[DESCRIPTION/SOUS-TITRE]

(: 4 :)
//enregistrement[description/annee < 1927
and description/mot-cle = "usine"]/@id-img

(: 5 :)
for $d in //DOCUMENT
where $d/DESCRIPTION/LANGUE = "Français-Chti"
return <res>{$d/DESCRIPTION/TITRE} {$d/DESCRIPTION/SOUS-TITRE}</res>

(: 6 :)
for $e in //enregistrement
return  if ($e/description/annee < 1930)
        then <image id="{data($e/@id-img)}" statut="avant 1930" />
        else <image id="{data($e/@id-img)}" statut="depuis 1930" />
```

```
(: 7 :)
for $e in //enregistrement,
  $d in //DOCUMENT
where $e/@id-img = $d/REFERENCES/REF-IMG
return <resultat>
  <image>{data($e/description/titre)}</image>
  <article>{data($d/DESCRIPTION/TITRE)}</article>
</resultat>
```

```
(: 8 :)
for $d in //DOCUMENT,
  $e in //enregistrement,
  $n in //notice
where $d/REFERENCES/REF-IMG = $e/@id-img
and $d/REFERENCES/REF-BAT = $n/@reference
and $e/description/titre = "Roubaix"
return <res>
  {$d/DESCRIPTION/TITRE}
  {$d/DESCRIPTION/AUTEUR}
  {$e/description/fichier}
  {$n/edifice}
</res>
```

```
(: 9 :)
for $n in //notice
group by $l := $n/localisation
let $c := avg($n/etages)
return <batiment localisation="{ $l }">
  <nb-etages>{ $c }</nb-etages>
</batiment>
```

```
(: 10 :)
for $d in //DOCUMENT
group by $t := $d/DESCRIPTION/TITRE
let $nbref := count($d/REFERENCES/REF-IMG) + count($d/REFERENCES/REF-BAT)
return <article titre="{ $t }">
  <nbref>{ $nbref }</nbref>
</article>
```