

INFO-H-509 : Technologies XML

TP 4 – XQuery

Professeur : Stijn Vansummeren

Assistant : Michaël Waumans

<http://cs.ulb.ac.be/public/teaching/infoh509>

2014 - 2015

DBLP est une base de données bibliographiques pour les sciences informatiques. Elle contient environ 1 million de références. Son contenu est disponible publiquement au format XML. Etant donné que son contenu représente plus de 400 Mb de données, seul un petit extrait sera utilisé pour les exercices suivants.

La collection DBLP respecte le format BibTex et contient les types de références suivants : article, proceedings, book, incollection, phdthesis, masters-thesis et www.

Les champs décrivant les types susmentionnés sont les suivants : author, editor, title, booktitle, pages, year, address, journal, volum, number, month, url, ee, cdrom, cite, publisher, note, crossref, isbn, series, school et chapter. Notez que tous les champs ne sont pas permis dans tous les types de références ; veuillez vous référer au fichier DTD pour plus d'informations à ce propos.

Définissez en XQuery les requêtes suivantes. Pour chaque requête, le format de sortie est spécifié.

Exercise 1.1

Donnez les types de publications dans le fichier.

```
<publications>
  <type>article</type>
  <type>book</type>
  <type>incollection</type>
  <type>inproceedings</type>
  <type>mastersthesis</type>
  <type>phdthesis</type>
  <type>proceedings</type>
</publications>
```

Exercise 1.2

Donnez le nombre de publications pour chaque type.

```
<publications>
  <reference type="article">
    <number>263</number>
  </reference>
  <reference type="book">
    <number>9</number>
  </reference>
```

```
...
</publications>
```

Exercise 1.3

Donnez la liste des noms des auteurs.

```
<authors>
  <name>A. A. Majid</name>
  <name>A. B. M. Shawkat Ali</name>
  <name>A. Bernoussi</name>
  ...
</authors>
```

Exercise 1.4

Donnez le nombre d'auteurs.

```
<nb_authors>1478</nb_authors>
```

Exercise 1.5

Donnez les noms des auteurs qui sont aussi éditeurs.

```
<authors_editors>
  <name>Ana Paiva</name>
  <name>Chris George</name>
  <name>Du Zhang</name>
  <name>Ed Dawson</name>
  ...
</authors_editors>
```

Exercise 1.6

Donnez le nombre de publications par auteur.

```
<author>
  <name>Ahmed Sameh</name>
  <number>3</number>
</author>
...
```

Exercise 1.7

Donnez les auteurs par ordre descendant du nombre de publications.

```
<author>
  <name>Christoph Meinel</name>
  <number>47</number>
</author>
<author>
  <name>Dieter Baum</name>
  <number>21</number>
</author>
...
```

Exercise 1.8

Donnez le(s) auteur(s) qui ont le plus grand nombre de publications.

```
<prolific_authors>
  <author>
    <name>Morshed U. Chowdhury</name>
    <number>5</number>
  </author>
</prolific_authors>
```

Exercise 1.9

Donnez pour chaque auteur, le nombre de types de publication.

```
<author>
  <name>Mazeyar E. Makoui</name>
  <pubtypes>1</pubtypes>
</author>
...
```

Exercise 1.10

Donnez pour chaque auteur le nombre total de publications et le nombre de publications par type.

```
<author>
  <name>A. B. M. Shawkat Ali</name>
  <total>3</total>
  <pubtype type="inproceedings">
    <number>3</number>
  </pubtype>
</author>

<author>
  <name>A. Bernoussi</name>
  <total>1</total>
  <pubtype type="article">
    <number>1</number>
  </pubtype>
</author>
...
```

Exercise 1.11

Donnez la liste des "proceedings" qui ont au moins un éditeur qui est aussi auteur d'au moins un article dans les "proceedings".

```
<proc_editor_author>
  <title>Proceedings of the International (...)</title>
  <title>Advanced Data Mining and Applications, (...)</title>
  <title>Proceedings of the 5th International (...)</title>
</proc_editor_author>
```

Exercise 1.12

Définissez une fonction qui imprime les références du type "article" au format HTML.

```
<html>
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
    <title>DBLP Articles</title>
  </head>
  <body>
    <h2>DBLP Articles</h2>
    <p>P. Berthon, C. B. Williams. Stages of e-democracy: (...)</p>
    ...
  </body>
</html>
```