

INFO-H-509 : Technologies XML

Projet 3 : XQuery

Introduction

Nous vous demandons d'écrire un ensemble de programmes XQuery permettant de réaliser les requêtes suivantes sur la base de données bibliographique DBLP introduite lors du Projet 2. Seulement un échantillon de la base de données de DBLP sera utilisée lors du projet.

Projet

L'objectif de ce projet est de réaliser un programme XQuery pour chacune des requêtes suivantes :

1. Pour chaque auteur, donnez le nombre de co-auteurs et le nombre de publications communes avec chacun d'entre eux. Veuillez utiliser le format suivant comme résultat de sortie :

```
<authors_coauthors>
  <author>
    <name>A. B. M. Shawkat Ali</name>
    <coauthors number="4">
      <coauthor>
        <name>M. Delowar Hossain</name>
        <nb_joint_pubs>1</nb_joint_pubs>
      </coauthor>
      ...
    </coauthors>
  </author>
  ...
</authors_coauthors>
```

2. Pour chaque *proceedings*, donnez son titre et le titre des articles contenus dans celui-ci. Veuillez utiliser le format suivant comme résultat de sortie :

```
<proceedings>
  <proc_title>6th Annual IEEE/ACIS International Conference (...)</proc_title>
  <title>Understanding Consumer Search Activity and Online (...)</title>
  <title>Approximate Element Computational Time for Domain (...)</title>
  <title>Towards a Table Driven XML QoS Aware Transmission Framework.</title>
  ...
</proceedings>
```

3. Nous définissons le graphe G des co-auteur de DBLP comme suit. G est un graphe non-dirigé dont l'ensemble des sommets est donné par l'ensemble des auteurs et tel qu'il existe une arête entre deux auteurs a et b dans G si et seulement si a et b ont écrit ensemble une publication. Nous définissons la distance entre deux auteurs a et b comme la longueur du plus court chemin entre a et b dans G. Par conséquent, deux auteurs ayant publié ensemble un même article ont une distance de un. De plus, si a et b n'ont pas publié ensemble, mais ont tout deux publié avec un troisième auteur c, alors la distance entre a et b est de deux.

Réalisez un programme XQuery qui calcule, pour chaque paire d'auteurs x et y distincts, la distance entre x et y en utilisant le format suivant pour le résultat de sortie :

```
<distances>
  <distance author1="Lizhu Zhou" author2="Dengfeng Zhang" distance="3"/>
  <distance author1="Lizhu Zhou" author2="Xuesong Yan" distance="2"/>
```

....
</distances>

Modalités

Le fichier (`dblp-extract.xml`) est disponible sur la page web du cours.

Tout comme lors des deux premiers projets, celui-ci contribuera à 2 points sur 20 de la note finale, l'examen écrit vaudra par conséquent 14 points sur 20.

Ce projet doit être réalisé par groupe de deux personnes. Il vous est demandé d'envoyer, par groupe, les noms des membres de votre groupe à Mr. Michaël Waumans (mwaumans@ulb.ac.be) pour le **28 Mars** au plus tard. Si vous ne parvenez pas à trouver un partenaire de travail, veuillez contacter Mr. Michaël Waumans par e-mail afin qu'il puisse vous indiquer un coéquipier.

Il vous est demandé de rendre, par groupe, un petit rapport (en Français ou en Anglais), contenant l'ensemble de vos hypothèses, documents XML d'exemples (lesquels doivent être valides pour le XSD) et document XSD de spécification. Chaque fichier devra être clairement documenté.

Le rapport ainsi que tous les documents requis doit être rendu à M. Michaël Waumans par email à l'adresse mwaumans@ulb.ac.be **au plus tard pour Lundi, 19 Mai, 2014**. (Veuillez mentionner comme sujet de votre e-mail : "INFO-H-509 - Projet 3")