UNIVERSITE LIBRE DE BRUXELLES 24 March 2021
Ecole Polytechnique

INFO-H509: XML and Web Technologies
Project 3: XQuery

Introduction

You are requested to write XQuery programs for the following queries against the DBLP bibliographical database excerpt introduced in Assignment 2.

Assignment

The goal of this assignment is to write an XQuery program for each of the following queries.

1. Give for each author the number of co-authors and the number of joint publications with each of them, using the following output format.

```
<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. For each proceedings give its title and the titles of articles appearing in it, using the following output format.

```
<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. Define the *co-author graph* G of the dblp-excerpt to be the undirected graph that has all authors as nodes, such that there is an edge between author a and author b in G if and only if a and b have written a publication together. Define the *distance* between two authors a and b to be the length of the shortest path between a and b in G. Hence, authors that have published together have distance 1. Moreover, if a and b have not published together but they have both published together with c, then the distance between a and b is two.

   Write an XQuery program that computes, for each pair of authors x and $y \neq x$ the distance between x and y using the following output format.

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

Important! Please ensure that all filename paths that you include in your queries are correctly capitalized. For example, if the file to load is named `dblp-excerpt.xml` then you should call `fn:doc('dblp-excerpt.xml')` and not `fn:doc('Dblp-Excerpt.xml')`. The former will work on all platforms, while the latter will work only on platforms whose file system is case-insensitive.

Your solutions will be tested on a *case-sensitive* platform. Hence, please be particulary attentive to all to your paths in your solutions. Any incoherence in the case of your paths that leads to an error on our testing environment will cost you points.

Submission

The required source file (`dblp-excerpt.xml`) may be found in Assignment 3 on UV.

This assignment contributes 5/20 to the overall grade.

This assignment is to be made in groups of two persons. You are asked to form the groups via the activity "Groups for Assignment 3" on UV.

You are asked to submit, per group, a small report (in English) containing all the hypotheses that you have made during your design, as well as the XQuery programs.
This report and all the required documents have to be uploaded as a zip file to "Assignment 3" on UV by **May 14, 2021**.