UNIVERSITE LIBRE DE BRUXELLES 26 March 2019
Ecole Polytechnique

INFO-H-509 : XML and Web Technologies
Project 3 : XQuery

Introduction

You are requested to write XQuery programs for the following queries against the DBLP bibliographical database introduced in Project Assignment 2. As with Project Assignment 2, only a small excerpt of the DBLP data will be used since the total DBLP database is more than 800 MB large (see the course's webpage for this excerpt).

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>
```

Modalities

The required source file (`dblp-extract.xml`) may be found on the course's website.

As with the first and second assignment, this third assignment contributes 2/20 to the overall grade . The written exam contributes the remaining 14/20 points.

This assignment is to be made in groups of two persons. You are asked to form the groups via the activity "Groupe selection for Assignment 3" on the *Université Virtuelle (UV)* by **April 2** at the latest. If one or both members of the group do not have access to the *UV*, please send an email with the names of the group members to Mr. Gilles Dejaegere (`Gilles.Dejaegere@ulb.ac.be`). If you cannot find a partner, please indicate so by also sending an email to Mr. Dejaegere, who will hook you up with a partner.

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 the activity "Submission Assignment 3" on the *UV* **no later than Tuesday, May 7, 2019**. Groups which do not have acces to the *UV* can send it by email to Mr. Gilles Dejaegere. (Please mention as subject of the email : "INFO-H-509 - Project 3" and add the lastnames of both team members in the name of each file sent)