

# XML Project 3 - Report

Artyom K.  
Ekeuh A.

May 2021

## 1 Introduction

This report presents details of explanation and hypothesis that lead to the implementation of 3 tasks of given assignment.

## 2 Code Structure

Each assignment is located in different xql documents

- task1.xql
- task2.xql
- task3.xql

The saxon tool is located in the project folder. *how\_to\_run* file has required commands needed to run the xql queries to fetch an output.

## 3 Architectural Decisions

### 3.1 Task 1

The first task requires to fetch all authors and their coauthors, also with the number of joint publications. To get data we start with select distinction values from document file by use of

```
for $authors in distinct-values(  
  doc("dblp-excerpt.xml")//author )
```

a for expression. Then we get author name by use of

```
data($authors)
```

To get co-authors we use again for expression with such structure:

```

for $coauthors in distinct-values(
  doc("dblp-excerpt.xml")/dblp/*[author=$authors]/author )
  where $authors ne $coauthors

```

where we indicate that we need authors that are different from what we already chosen.

Finally we get the names of co-authors and their number of joint publications by use of the following part of code:

```

<name> { data($coauthors) } </name>
<nb_joint_pubs>
  { count(distinct-values(doc("dblp-excerpt.xml")/
    dblp/*[author=$authors]/author[.=$coauthors] )) }
</nb_joint_pubs>

```

and we finally enclose the fetched data with `jauthors_coauthors` tag.

### 3.2 Task 2

In the second task we need to get proceedings with all in proceedings included. The solution is straightforward and we start with iterating through proceedings by this way:

```

for $proceeding in
  doc("dblp-excerpt.xml")/dblp/proceedings

```

By doing this we get proceedings, but still we need to get inproceedings that connected to them. For doing that we use:

```

for $inproceeding in
  doc("dblp-excerpt.xml")/dblp/inproceedings
  where $proceeding/@key = $inproceeding/crossref

```

As you see, we fetch in-proceedings and we need connect them to our proceedings. For doing that we join them by common data that is `@key` in proceedings and `./crossref` in in-proceedings.

### 3.3 Task 3

In the third task we need to get a distance between authors, in terms of cooperation between each of them. If author has been written book or whatever with another author, then the connection is one, if they had some intermediate author (c) then 2, and so on. We decided to limit our search up to 3 connections, since it takes a lot of data to process.

First of all we get select some author and his co-authors by following command:

```

for $select_author in distinct-values(
doc("dblp-excerpt.xml")//author ), $coauthor in
distinct-values(doc("dblp-excerpt.xml")/dblp/*
[author=$select_author]/author)
where data($select_author) != data($coauthor)

```

By doing that we already get direct connections between some authors. We need to proceed further and get connections that have intermediate authors. We are fetching data by this command:

```

for $other_coauthor in
distinct-values(doc("dblp-excerpt.xml")/dblp/*
[author=$coauthor]/author)
where $select_author ne $other_coauthor

return

for $compare_author in distinct-values(
doc("dblp-excerpt.xml")/
dblp/*[author=$select_author]/author)
where not($compare_author = $select_author) and
$compare_author != $other_coauthor

```

The important thing here is that we find *other\_author* and then we are taking coauthors of him here. The result we get are intermediate connections. We are doing the same procedure for deeper connection (through c and d authors correspondingly). As a result we get a big output document with all connections we specified.