Exam

Web Data Models
Université Paris-Saclay, M2 Data&Knowledge

November 6th, 2017

This is the final exam for the Web Data Models class, which will determine 50% of your grade for this class. The duration of the exam is two and a half hours. This exam subject consists of 4 exercises and has 4 pages.

You are allowed one A4 sheet of personal notes, written by hand. If you use such personal notes, you must hand them in along with your answers. You may not use any other written material.

The exam is strictly personal: any communication or influence between students, or use of outside help, is prohibited. Any violation of the rules may result in a grade of 0 and/or disciplinary action.

Exercise 1 (2 points)

Consider the following XML document, `ex1.xml`:

```xml
<directory>
  <person>
    <name>
      <first>Pierre</first>
      <last>Duchamel</last>
    </name>
    <salary>35000</salary>
    <location>
      <street>rue des Bureaux</street>
      <city>Paris</city>
      <office>N/11</office>
    </location>
  </person>
  <person>
    <name>
      <last>Valjean</last>
    </name>
    <location>
      <street>rue de Paris</street>
      <city>Versailles</city>
      <tel>06543311</tel>
    </location>
  </person>
  <person>
    <name>
```

```
    <first>Valerie</first>
    <last>Bovary</last>
  </name>
  <salary>55000</salary>
  <location>
    <street>rue Aristide Briand</street>
    <city>Antony</city>
    <tel>01347619</tel>
    <office>S/109</office>
  </location>
 </person>
</directory>
```

Question 1. Is the document in `ex1.xml` *well-formed*? If it not well-formed, show the parts of the XML that are not well-formed, and give a corrected version *which maintains the same information.*

Question 2. For the corrected version of `ex1.xml`, draw its corresponding document tree.

Exercise 2 (6 points)

In this exercise, we will work with the following two grammars and the document `ex1.xml` or its corrected version from Exercise 1:

G_1	G_2
	$Dir \rightarrow directory[(Staff|Client)+]$
$Dir \rightarrow directory[Person*]$	$Staff \rightarrow person[Name.Salary.LocationStaff]$
$Person \rightarrow person[Name.Salary.Location]$	$Client \rightarrow person[Name.LocationClient]$
$Name \rightarrow name[First?.Last]$	$Name \rightarrow name[First?.Last]$
$First \rightarrow first[PCData]$	$First \rightarrow first[PCData]$
$Last \rightarrow last[PCData]$	$Last \rightarrow last[PCData]$
$Salary \rightarrow salary[PCData]$	$Salary \rightarrow salary[PCData]$
$Location \rightarrow location[Street.City.Tel?.Office]$	$LocationStaff \rightarrow location[Street.City.Tel?.Office]$
$Street \rightarrow street[PCData]$	$LocationClient \rightarrow location[Street.City.Tel]$
$City \rightarrow city[PCData]$	$Street \rightarrow street[PCData]$
$Tel \rightarrow tel[PCData]$	$City \rightarrow city[PCData]$
$Office \rightarrow office[PCData]$	$Tel \rightarrow tel[PCData]$
	$Office \rightarrow office[PCData]$

Question 1. Is G_1 an LTG or a STTG? What about G_2? Justify.

Question 2. Give the tree automaton A_1 corresponding to G_1. Detail the run of the automaton A_1 on `ex1.xml` or its corrected version. Is the run *accepting*?

Question 3. Give the tree automaton A_2 corresponding to G_2. Detail the run of the automaton A_2 on `ex1.xml` or its corrected version. Is the run *accepting*?

Question 4. Considering your answers to Questions 1, 2, and 3, what language would you use to validate `ex1.xml` or its corrected version, DTD or XML Schema? Justify why.

Exercise 3 (6 points)

Consider the following XML document, `ex3.xml`:

```
<a>
    <a>
        <a>
            <b />
            <b>
                <c />
            </b>
        </a>
        <b />
    </a>
    <c>
        <a>
            <b />
        </a>
    </c>
</a>
```

Question 1 (3 points). Consider the query $Q_1 = //a/a/b$. What fragment of XPath does it belong to, simple or core? Detail the run of a correct evaluation algorithm on `ex3.xml`, by showing the steps of the algorithm, the data structures used, and the results.

Question 2 (3 points). Consider the query $Q_2 = //a[following::c]$. What fragment of XPath does it belong to, simple or core? Detail the run of a correct evaluation algorithm on `ex3.xml`, by showing the steps of the algorithm, the data structures used, and the results.

Exercise 4 (6 points)

Consider the XML document below, `ex4.xml`:

```
<a>
    <a>
        <b>
            <a />
        </b>
        <c>
            <d />
            <d />
        </c>
        <d>
            <c>
                <a />
                <a />
            </c>
        </d>
    </a>
    <b>
        <b>
            <a />
```

```
            </b>
            <b>
                <a />
                <d />
            </b>
        </b>
        <b />
    </a>
```

Question 1 In the above document, the *context* node is the node labelled `author`, uniquely identified by the XPath query /a/a.

Consider the following XML axes:

1. following

2. preceding

3. ancestor

4. preceding-sibling

5. parent

For each axis, give the list of nodes as their *pre-order node ID*, considering that the root of the document has ID 0.

Question 2 Consider again the XML document above. For each of the XPath queries below, give the list of the XML document nodes satisfying the query, as a sequence of node IDs, where each ID is the *pre-order* ID, considering that the root of the document is the node having ID 0:

1. //a/b

2. //b[d]

3. //b[a and not d]

4. /a//d/preceding::b

5. /a//d[preceding::b]

Question 3 For each pair P, Q of XPath queries below, indicate whether $P \subseteq_0 Q$, i.e., P is 0-contained in Q. Give a justification why, or show a counter-example document.

1. P=/a[//c[b]/d/b[e]/c and Q=/a/b[//e]/c

2. P=/c[b]/a/d and Q=/c[b]//a[d]

3. P=/a[//b]/c/b and Q=/a[/c[/d]/b]//b