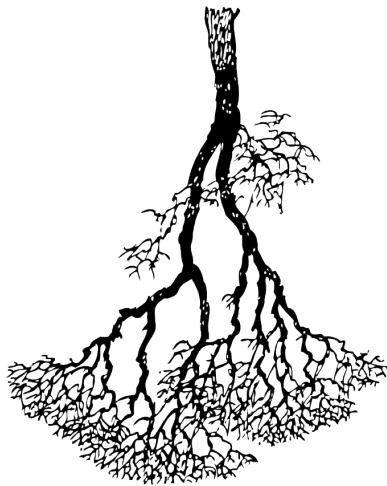




DEPARTEMENT INFORMATIQUE  
DE LA FACULTE DES SCIENCES

Quentin Yeché (21520370), Yanis Allouch (21708237)

## Rapport du TP N°1: XML et DTDs



**HMIN103 — Données du web**

Référent: Federico Ulliana et Pierre Pompidor

2020

## Table des matières

<b>1</b>	<b>Exercice</b>	<b>3</b>
1.1	Dès schémas aux données . . . . .	3
1.2	Dès données aux schémas . . . . .	4
1.3	Modélisation : à quoi ressemble un Tweet ? . . . . .	5
<b>2</b>	<b>Annexe</b>	<b>6</b>

## Introduction

L'analyse et le traitement des données consistent à les étudier afin d'en extraire des structures capable des les stocker ainsi que de les manipuler de la façon la plus optimal.

L'objectif de ce premier TP est de manipuler et traiter la structure de donnée XML et DTD à partir d'exemple de tweet.

Le TP se déroule sur papier.

Durant ce TP, nous appliquerons les règles de constructions de XML et de DTD abordées dans le cours référencé [DdW1-1-XML](#) et [DdW1-2-DTD](#).

## 1 Exercice

### 1.1 Dès schémas aux données

```
1 <!--Exo 1 Arbre 1 version 1-->
2 <presse>
3   <journal>
4     <nom>Times</nom>
5     <directeur>
6       <nom>Williams</nom>
7       <prenom>William</prenom>
8     </directeur>
9   </journal>
10  <journalistes>
11    <journaliste anonymisation="non" idj="j01">
12      <nom>Poe</nom>
13      <prenom>Edgar</prenom>
14    </journaliste>
15  </journalistes>
16 </presse>
```

```
1 <!--Exo 1 Arbre 1 version 2-->
2 <presse>
3   <journal>
4     <nom>Times</nom>
5     <directeur>
6       <nom>Williams</nom>
7       <prenom>William</prenom>
8     </directeur>
9   </journal>
10  <journalistes>
11    <journaliste anonymisation="oui" idj="j01">
12      <pseudonyme>Anatole France</pseudonyme>
13    </journaliste>
14  </journalistes>
15 </presse>
```

```
1 <!--Exo 1 Arbre 2 version 1-->
2 <batiment>
3   <etage>
4     <description>foobar</description>
5     <bureau>
6       <code>foobar</code>
7     </bureau>
8     <salle>
9       <nombrePlaces>foobar</nombrePlaces>
10    </salle>
11  </etage>
12 </batiment>
```

```

1 <!--Exo 1 Arbre 2 version 2-->
2 <batiment>
3   <etage>
4     <description>foobar</description>
5     <bureau>
6       <code>foobar</code>
7       <personne>william</personne>
8     </bureau>
9     <salle>
10      <nombrePlaces>foobar</nombrePlaces>
11    </salle>
12  </etage>
13 </batiment>

```

## 1.2 Dès données aux schémas

La liste étant assez longue et ne nécessitant pas plus d'explication, vous retrouverez le reste des codes en annexe.

```

1 <!--XML1 Version 1-->
2 <!DOCTYPE F [
3   <!ELEMENT F (C*,B*)>
4   <!ELEMENT C (A|D)>
5   <!ELEMENT B (A|E)>
6   <!ELEMENT A EMPTY>
7   <!ELEMENT D EMPTY>
8   <!ELEMENT E EMPTY>
9 ]>

```

```

1 <!--XML1 Version 2-->
2 <!DOCTYPE F [
3   <!ELEMENT F (C,B,B)>
4   <!ELEMENT C (A|D)>
5   <!ELEMENT B (A|E)>
6   <!ELEMENT A EMPTY>
7   <!ELEMENT D EMPTY>
8   <!ELEMENT E EMPTY>
9 ]>

```

### 1.3 Modélisation : à quoi ressemble un Tweet ?

```
1 <!DOCTYPE tweet[
2   <!--!ELEMENT tweet (metadata?,body)-->
3   <!--!ATTLIST tweet idT ID #REQUIRED-->
4
5   <!--!ELEMENT metadata (media?,location?,userConfig?)-->
6   <!--!ELEMENT media (url+)-->
7   <!--!ELEMENT url (#PCDATA)-->
8
9   <!--!ELEMENT location (gps?,city?,country?)-->
10  <!--!ELEMENT gps (lat,long)-->
11  <!--!ELEMENT lat (#PCDATA)-->
12  <!--!ELEMENT long (#PCDATA)-->
13  <!--!ELEMENT city (#PCDATA)-->
14  <!--!ELEMENT country (#PCDATA)-->
15
16  <!--!ELEMENT userConfig (version)-->
17  <!--!ATTLIST userConfig osType (Android|iOS|Windows|Linux|
18    unknown|other) "unknown"-->
19  <!--!ELEMENT version (#PCDATA)-->
20
21  <!--!ELEMENT body (formatting,language,retweets,author,content)-->
22  <!--!ATTLIST body idOriginalTweet IDREF #IMPLIED-->
23
24  <!--!ELEMENT formatting (fontsize,fontcolor,font)-->
25  <!--!ELEMENT fontsize (#PCDATA)-->
26  <!--!ELEMENT fontcolor (#PCDATA)-->
27  <!--!ELEMENT font (#PCDATA)-->
28
29  <!--!ELEMENT language (#PCDATA)-->
30  <!--!ELEMENT retweets (#PCDATA)-->
31
32  <!--!ELEMENT author (name,url)-->
33  <!--!ATTLIST idAuthor IDREF #REQUIRED-->
34  <!--!ELEMENT name (#PCDATA)-->
35
36  <!--!ELEMENT content ((contentElement)*)-->
37  <!--!ELEMENT contentElement (#PCDATA|url|hashtag|userref)-->
38
39  <!--!ELEMENT hashtag (#PCDATA)-->
40  <!--!ATTLIST idHashtag IDREF #REQUIRED-->
41
42  <!--!ELEMENT userref (#PCDATA)-->
43  <!--!ATTLIST idUser IDREF #REQUIRED-->
44 ]>
```

## 2 Annexe

```
1 <!-- XML2 Version 1 -->
2 <!DOCTYPE D [
3 <!ELEMENT D (C*|E*|B*)>
4 <!ELEMENT C EMPTY>
5 <!ELEMENT E EMPTY>
6 <!ELEMENT B EMPTY>
7 ]>
```

```
1 <!--XML2 Version 2 -->
2 <!DOCTYPE D [
3 <!ELEMENT D (C*|E*|B*)>
4 <!ELEMENT C EMPTY>
5 <!ELEMENT E EMPTY>
6 <!ELEMENT B EMPTY>
7 ]>
```

```
1 <!--XML3 Version 1-->
2 <!DOCTYPE A [
3 <!ELEMENT A (C,C,E,E,E,C)>
4 <!ELEMENT C EMPTY>
5 <!ELEMENT E EMPTY>
6 ]>
```

```
1 <!-- XML4 Version 1 -->
2 <!DOCTYPE B [
3 <!ELEMENT B (C*|E*|B*|D*)>
4 <!ELEMENT C EMPTY>
5 <!ELEMENT E EMPTY>
6 <!ELEMENT B EMPTY>
7 <!ELEMENT D EMPTY>
8 ]>
```

```
1 <!DOCTYPE B [
2 <!-- XML4 Version 2 -->
3 <!ELEMENT B (C*|E*|B*|D*)>
4 <!ELEMENT C EMPTY>
5 <!ELEMENT E EMPTY>
6 <!ELEMENT B EMPTY>
7 <!ELEMENT D EMPTY>
8 ]>
```

```

1 <!--XML3 Version 2-->
2 <!DOCTYPE A [
3   <|ELEMENT A ((C|E)*)>
4   <|ELEMENT C EMPTY>
5   <|ELEMENT E EMPTY>
6 ]>

```

```

1 <!--XML5 Version 1-->
2 <!DOCTYPE A [
3   <|ELEMENT A EMPTY>
4 ]>

```

```

1 <!--XML5 Version 2-->
2 <!DOCTYPE A [
3   <|ELEMENT A ANY>
4 ]>

```

```

1 <!DOCTYPE A [
2   <!-- XML6 Version 1 -->
3   <|ELEMENT A (C+|E*|B*|D*)>
4   <|ELEMENT C (#PCDATA)>
5   <|ATTLIST C id IDREF #REQUIRED>
6   <|ATTLIST C id ID #REQUIRED>
7   <|ELEMENT C EMPTY>
8 ]>

```

```

1 <!DOCTYPE A [
2   <!-- XML6 Version 2 -->
3   <|ELEMENT A (C+)>
4   <|ELEMENT C (#PCDATA)>
5   <|ATTLIST C id IDREF #REQUIRED>
6   <|ATTLIST C id ID #REQUIRED>
7   <|ELEMENT C EMPTY>

```