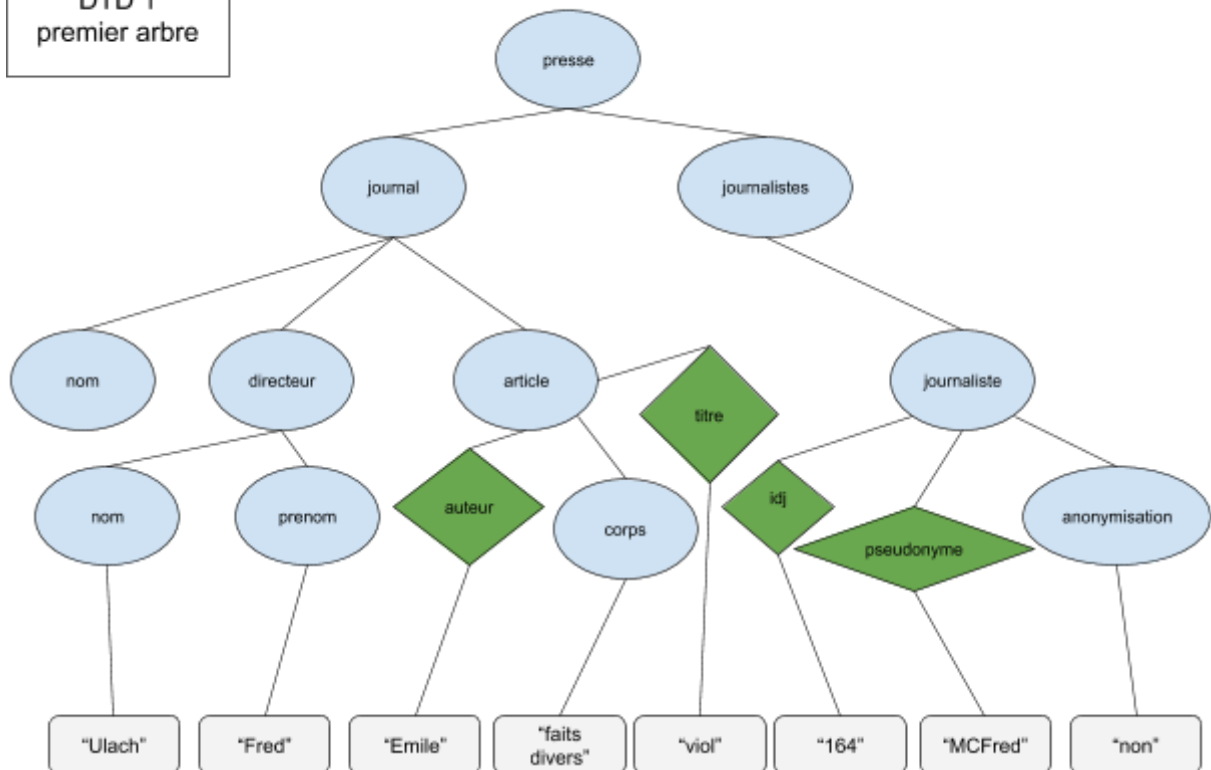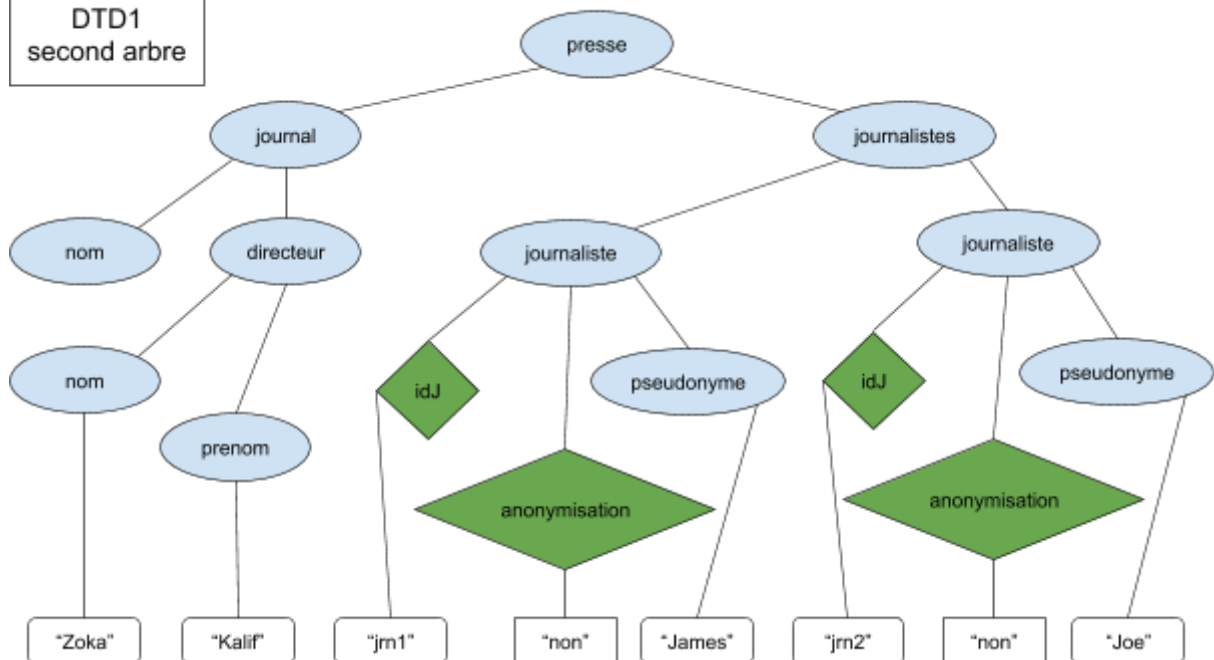Emile Elie Youssef
Terki Mohamed Adel

HMIN 103 : Données du WEB
TD1 : XML & DTDs (en binôme)

1) Des schémas aux données :

DTD 1
premier arbre

DTD1
second arbre

DTD 2
premier arbre

batiment

étage

description

"Service administratif"

bureau

code

"SA140"

DTD 2
second arbre

batiment

étage

description

"deuxième étage"

salle

nombrePlaces

"44"

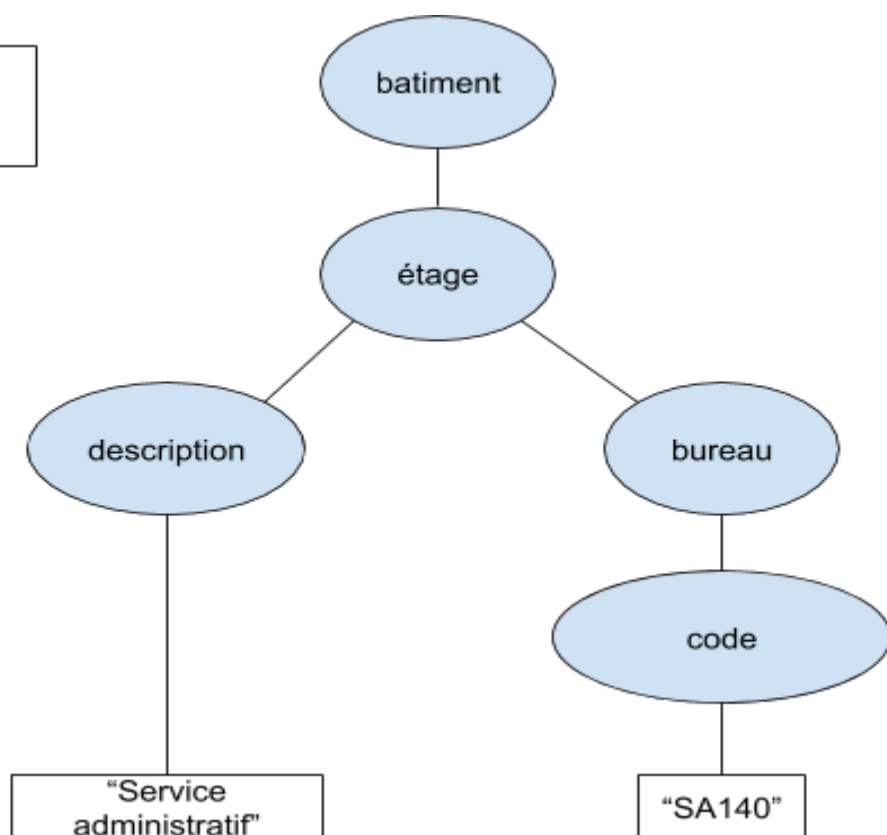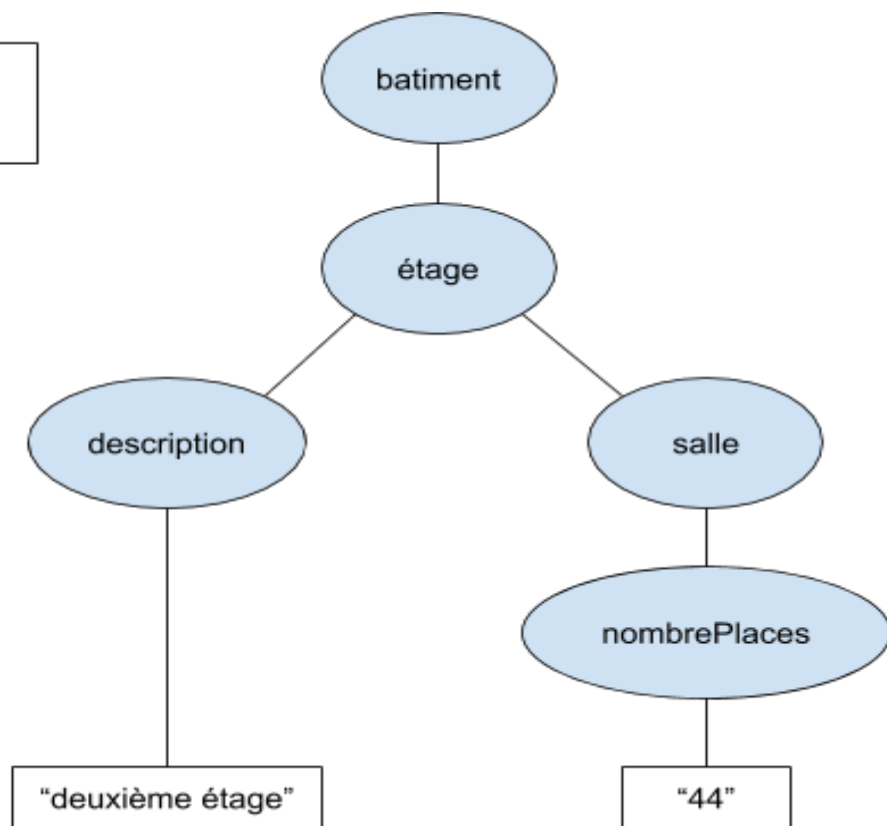**2) Des données aux schémas**

XML 1:

1)   <! DOCTYPE A[
            <! ELEMENT A (C,B)+>
            <! ELEMENT B (A*,E*)>
            <! ELEMENT C (A*,D*)>
            <! ELEMENT D (#EMPTY)>
            <! ELEMENT E (#EMPTY)>
        ]>

2)   <! DOCTYPE A[
            <! ELEMENT A (C,B)*>
            <! ELEMENT B (A|E)>
            <! ELEMENT C (A|D)>
            <! ELEMENT D (#EMPTY)>
            <! ELEMENT E (#EMPTY)>
        ]>

XML 2:

1)   <!DOCTYPE B [
            <!ELEMENT B (B, C+, E*)>
            <!ELEMENT C (#EMPTY)>
            <!ELEMENT E (#EMPTY)>
        ]>

2)   <!DOCTYPE B [
            <!ELEMENT B (B, C+, E?)>
            <!ELEMENT C (#EMPTY)>
            <!ELEMENT E (#EMPTY)>
        ]>

XML 3:

1)   <! DOCTYPE A[
            <! ELEMENT A (C,B)+>
            <! ELEMENT B (#EMPTY)>
            <! ELEMENT C (#EMPTY)>
        ]>

2)   <! DOCTYPE A[
            <! ELEMENT A (C,B)*>
            <! ELEMENT B (#EMPTY)>
            <! ELEMENT C (#EMPTY)>
        ]>

XML 4 :

1)   <!DOCTYPE B [
            <!ELEMENT B (C, D,E)+>
            <!ELEMENT C (#EMPTY)>
            <!ELEMENT D (#EMPTY)>
            <!ELEMENT E (#EMPTY)>
        ]>

2)   <!DOCTYPE B [
            <!ELEMENT B (C+, D,E)*>
            <!ELEMENT C (#EMPTY)>
            <!ELEMENT D (#EMPTY)>
            <!ELEMENT E (#EMPTY)>
        ]>

XML 5:

1)   <! DOCTYPE A[
            <! ELEMENT A (#EMPTY)>
        ]>

2)   <! DOCTYPE A[
            <! ELEMENT A (A)*>
        ]>

XML 6:

1)   <! DOCTYPE A[
            <! ELEMENT C (#PCDATA)>
            <! ATTLIST C id   CDATA #REQUIRED>
            <! ATTLIST C friend CDATA #REQUIRED>
        ]>

2)   <! DOCTYPE A [
            <! ELEMENT C (#PCDATA)>
            <! ATTLIST C id   CDATA #IMPLIED>
            <! ATTLIST C friend CDATA #IMPLIED>
        ]>

3) **Modélisation : à quoi ressemble un Tweet ?**

```
<!DOCTYPE tweet [
      <!ELEMENT tweet (date, message?, gps, user, url_media*)>
      <!ATTLIST tweet idT ID #REQUIRED>

      <!ELEMENT date (seconds, GMT)>
      <!ELEMENT seconds  (#PCDATA)>
      <!ELEMENT GMT   (#PCDATA)>

      <!ELEMENT message (text, ref_tag*, ref_user*)>
      <!ATTLIST message length  CDATA #IMPLIED>
      <!ATTLIST message type    CDATA #IMPLIED>
      <!ATTLIST message color   CDATA #IMPLIED>
      <!ATTLIST message lang    CDATA #IMPLIED>
      <!ELEMENT text       (#PCDATA)>
      <!ELEMENT ref_tag   (#PCDATA)>
      <!ELEMENT ref_user (#PCDATA)>

      <!ELEMENT gps (latitude, longitude, altitude)?>
      <!ELEMENT latitude    (#PCDATA)>
      <!ELEMENT longitude (#PCDATA)>
      <!ELEMENT altitude    (#PCDATA)>

      <!ELEMENT user (name, url_avatar, follow_count, sub_count, description?)>
      <!ATTLIST user idU       IDREF #REQUIRED>
      <!ATTLIST user city      CDATA #REQUIRED )>
      <!ATTLIST user country CDATA #REQUIRED)>
      <!ATTLIST user OS       CDATA #REQUIRED)>
      <!ELEMENT name          (#PCDATA)>
      <!ELEMENT url_avatar     (#PCDATA)>
      <!ELEMENT follow_count  (#PCDATA)>
      <!ELEMENT sub_count     (#PCDATA)>
      <!ELEMENT description    (#PCDATA)>

      <!ELEMENT url_media (#PCDATA)>
]>
```

4) Instance de données

```xml
<tweet idT="1">
  <date>
    <seconds>40</seconds>
    <GMT>1</GMT>
  </date>
  <message length="140" type="text" color="black" lang="en">
    <text><ref_user>example</ref_user> absolutely smashed it at
<ref_tag>mtvlivelockdown</ref_tag> ! Catch him at the official
<ref_user>clubmtvuk</ref_user> after party tonight @10 pm</text>
  </message>
  <gps>
    <latitude>0</latitude>
    <longitude>0</longitude>
    <altitude>0</altitude>
  </gps>
  <user idU="1" city="Montpellier" country="France" OS="Windows">
    <name>MTV Music</name>
    <url_avatar>null</url_avatar>
    <follow_count>3000</follow_count>
    <sub_count>12042</sub_count>
    <description>MTV Music UK</description>
  </user>
  <url_media></url_media>
</tweet>
```

5) Synthèse des modèles

LEUR BINÔME

```
<tweet id="9348398423" lang="english">
  <message font-size="12" font-type="normal" color="standard">
  .@example absolutely smashed it at #mtvlivelockdown!
Catch him at the official @clubmtvuk after party tonight @ 10pm
  </message>
  <timestamp timezone="UTC">1536919024</timestamp>
  <author id="8821189298219">
    <name>MTV Music</name>
    <url>//twitter.com/MTVMusic</url>
    <description>music / pop stars.</description>
    <picture format="jpg">RVpLLjMyMzkuLi4=...</picture>
    <follower-count>666000</follower-count>
    <subscription-count>5952</subscription-count>
  </author>
</tweet>
```

NOTRE BINÔME

```
<tweet idT="1">
 <date>
  <seconds>40</seconds>
  <GMT>1</GMT>
 </date>
 <message length="140" type="text" color="black" lang="en">
  <text><ref_user>example</ref_user> absolutely smashed it at
    <ref_tag>mtvlivelockdown</ref_tag> ! Catch him at the official
<ref_user>clubmtvuk</ref_user> after party tonight @10 pm</text>
 </message>
 <gps>
  <latitude>0</latitude>
  <longitude>0</longitude>
  <altitude>0</altitude>
 </gps>
 <user idU="1" city="Montpellier" country="France" OS="Windows">
  <name>MTV Music</name>
  <url_avatar>null</url_avatar>
  <follow_count>3000</follow_count>
  <sub_count>12042</sub_count>
  <description>MTV Music UK</description>
 </user>
 <url_media></url_media>

</tweet>
```

Différence entre les modèles :

__ : utilisation des balises de référencement dans la partie texte afin de les démarquer
__ : stockage de la langue en attribut dans des éléments distincts : message et tweet
__ : stockage de l'OS en tant qu'attribut de l'utilisateur contre un stockage en élément du tweet

**Unification des modèles :**

```xml
<tweet idT="1" lang="en">
  <timestamp timezone="UTC">
      123059O4
  </timestamp>
  <message length="140" type="text" color="black">
    <text><ref_user>example</ref_user> absolutely smashed it at
ref_tag>mtvlivelockdown</ref_tag> ! Catch him at the official
<ref_user>clubmtvuk</ref_user> after party tonight @10 pm</text>
  </message>
  <gps>
    <latitude>0</latitude>
    <longitude>0</longitude>
    <altitude>0</altitude>
  </gps>
  <user idU="1" city="Montpellier" country="France" OS="Windows">
    <name>MTV Music</name>
    <url_avatar>null</url_avatar>
    <follow_count>3000</follow_count>
    <sub_count>12042</sub_count>
    <description>MTV Music UK</description>
  </user>
  <url_media></url_media>
</tweet>
```

6) **Cas particuliers des DTDs**

DTD 1:

```
<!DOCTYPE B>
<B>
   <A><D>something</D></A>
   <C>something<D>another thing</D></C>
</B>
```

DTD 2:

```
<!DOCTYPE EMPTY>
<EMPTY/>
```

DTD 3:

```
<!DOCTYPE A>
<A>
   <B><A/></B>
</A>
```

DTD 4:

```
<!DOCTYPE C>
<C><C/><C/></C>
<C><C/></C>
```

DTD 5:

```
<!DOCTYPE C>
<C>
      <C/><EMPTY/>
</C>
```

DTD 6:

L'élément B n'est pas déclaré, il n'est pas possible de donner un document XML valide.