

Données du Web - TD2 1

Basil Dalié - Yannis Naidja

Octobre 2019

Chapitre 1

XPath : Films

1.1 Les titres des films

Donner les requêtes suivantes en XPath. (Voir DTD en annexe)

1.2 Les titres des films

```
1 /FILMS/FILM/TITRE/text()
```

1.3 Les titres des films parus en 1990

```
1 /FILMS/FILM/ANNEE[text() = '1990']/parent::FILM/TITRE/text()
```

1.4 Le resume d'Alien

```
1 /FILMS/FILM/TITRE[text() = 'Alien']/parent::FILM/RESUME/text()
```

1.5 Quel est le dernier film du document ?

```
1 /FILMS/FILM[last()]
```

1.6 Quel est le dernier film du document paru en 1990 ?

```
1 (/FILMS/FILM/ANNEE[text() = '1990'])[last()]/parent::FILM
```

1.7 Les titres des films qui ont un résumé

```
1 /FILMS/FILM/RESUME[text() != '']/parent::FILM/TITRE/text()
```

1.8 Les titres des films dont l'élément résumé n'est pas présent.

```
1 /FILMS/FILM[not(RESUME)]/TITRE/text()
```

1.9 Donnez les noms des acteurs qui ont joué dans Vertigo.

```
1 /FILMS/ARTISTE[@id  
  ↳ =/FILMS/FILM[TITRE='Vertigo']/ROLES/ROLE/@idref]/ACTNOM/text()
```

1.10 Qui a mis en scène Vertigo ?

```
1  
  ↳ /FILMS/ARTISTE[@id=/FILMS/FILM[TITRE='Vertigo']/MES/@idref]/ACTNOM/text()
```

1.11 Donnez tout les films du directeur de Vertigo.

```
1  
  ↳ /FILMS/FILM/MES[@idref=/FILMS/FILM[TITRE='Vertigo']/MES/@idref]/preceding-sibling::TI
```

1.12 Donnez les titres des films qui contiennent la lettre "V" (utiliser la fonctioncontains()).

```
1 /FILMS/FILM/TITRE[contains(text(), 'V')]/text()
```

1.13 Les titres des films où l'acteur Bruce Willis a joué.

```
1  
  ↳ /FILMS/FILM/ROLES/ROLE[@idref=/descendant::ARTISTE[ACTNOM/text()='Willis']  
  ↳ and  
  ↳ ACTPNOM/text()='Bruce']/number(@id)]/parent::ROLES/parent::FILM/TITRE
```

1.14 Quel rôle joue Harvey Keitel dans Reservoir dogs ?

```
1  /FILMS/FILM[TITRE = 'Reservoir dogs']/ROLES/ROLE[@idref =  
  ↪ /FILMS/ARTISTE[ACTNOM = 'Keitel' and ACTPNOM =  
  ↪ 'Harvey']/@id]/text()
```

1.15 Qui a joué avec Harvey Keitel dans Reservoir dogs ?

```
1  /FILMS/ARTISTE[@id=/FILMS/FILM[TITRE/text() = "Reservoir  
  ↪ dogs"]/ROLES/ROLE/@idref and ACTNOM/text() != 'Keitel' and  
  ↪ ACTPNOM/text() != 'Harvey']
```

1.16 Donnez les nœuds qui ont exactement trois descendants (utiliser la fonctioncount()).

```
1  descendant::*[count(*)=3]
```

1.17 Donnez les nœuds dont le nom contient la chaîne "TI" (utiliser la fonctionname()).

```
1  descendant::*[contains(name(), 'TI')]
```

1.18 Quel est le titre du film qui précède immédiatement Shining (dans l'ordre du document) ?

```
1  /FILMS/FILM[TITRE =  
  ↪ 'Shining']/preceding-sibling::FILM[1]/TITRE/text()
```

Chapitre 2

XPath : Recettes

Exprimer en XPath les interrogations suivantes. Voir document XML en annexe.

2.1 Le nom complet de toutes les recettes

```
1 /recettes/recette/@nom
```

2.2 Les ingrédients de la recette dont le nom court est "Chiffonnade" ;

```
1 /recettes/recette[@nomCourt =  
  ↳ "Chiffonnade"]/materiel/ingredient/text()
```

2.3 Le nom complet des recettes utilisant du "persil" ;

```
1 /recettes/recette/materiel/ingredient[contains(text(),  
  ↳ "persil")]/parent::materiel/parent::recette/@nom
```

2.4 (Sans utiliser l'axechild) Le nom complet des recettes utilisant du "persil"

```
1 /descendant::ingredient[contains(text(),  
  ↳ "persil")]/ancestor::recette/@nom
```

2.5 Le nom complet des recettes ayant plus de deux ingrédients, et contenant des oeuf

```
1 /recettes/recette[count(materiel/ingredient) > 2 and  
  → materiel/ingredient[contains(text(), "oeuf")]]/@nom
```

2.6 (Sans utiliser la fonctioncount()) Le nom complet des recettes ayant plus de deux ingrédients, et contenant l'ingrédient "huile

```
1 /recettes/recette/materiel/ingredient[last() > 2 and  
  → contains(text(),  
  → "huile")]/parent::materiel/parent::recette/@nom
```

2.7 La dernière recette du document

```
1 /recettes/recette[last()]/@nom
```

Chapitre 3

XPath : Trains

Exprimer en XPath les interrogations suivantes. Voir document XML en annexe.

3.1 Le numéro des trains qui possèdent une voiture-bar

```
1 /gare/train/voiture/bar/parent::voiture/parent::train/@numero)
```

3.2 Le nom des usages ayant effectué au moins une réservation ;

```
1 /gare/usager[@id = /gare/train/voiture/resa/@id]/@nom
```

(Extra) Est-il possible d'exprimer en XPath les requêtes suivantes ?

3.3 Le numéro des trains dont au moins 2 places sont réservées :

```
1 /gare/train/voiture[count(resa) >= 2]/parent::train/@numero
```

3.4 Le nom des personnes ayant réservé exactement deux fois.

Impossible, car en XPath on ne peut pas compter le nombre d'occurrence d'un attribut

3.5 Les usagers n'ayant effectué aucune réservation

```
1 /gare/usager[not(@id = /gare/train/voiture/resa/@id)]
```


Chapitre 4

XPath : Tweets

Reprenez votre DTD pour les Tweets, et créez un document XML valide contenant au moins 3 utilisateurs et 5 tweets. Attention : pour le bon déroulement de l'exercice, il sera peut être nécessaire d'apporter des légères modifications à votre DTD afin qu'il soit possible d'interroger vos données !

Donner les requêtes XPath correspondants aux expressions suivantes et évaluer ces expressions dans le document XML créé pour les Tweets.

Voir DTD et document XML exemple en annexe.

4.1 Les noms des auteurs des tweets.

```
1 /tweeter/users/user[@id =  
  ↳ /tweeter/tweets/tweet/@author_ref]/concat(first_name, ' ',  
  ↳ last_name)
```

4.2 Les tweets de l'utilisateur dont l'id est "u41".

```
1 /tweeter/users/user[@id =  
  ↳ /tweeter/tweets/tweet/@author_ref]/concat(first_name, ' ',  
  ↳ last_name)
```

4.3 Les tweets contenant l'hashtag "#I;3XML".

```
1 /tweeter/tweets/tweet/body/hashtag[contains(text(),  
  ↳ 'I<3XML')]/ancestor::tweet
```

4.4 Le tweet le plus recent.

```
1 /tweeter/tweets/tweet/header/date[text() =  
  ↳ max(/tweeter/tweets/tweet/header/date)]/parent::header/parent::tweet
```

4.5 Les tweet sans hashtags.

```
1 /tweeter/tweets/tweet/body[count(hashtag) = 0]/parent::tweet
```

4.6 Les retweets du tweet dont l'id est "t42".

```
1 /tweeter/tweets/tweet[@id = /tweeter/tweets/tweet[@id =  
  ↳ "T42"]/header/retweets/retweet/@ref]
```

4.7 Les utilisateurs ayant répondu au tweet dont l'id est "t42".

```
1 /tweeter/users/user[@id =  
  ↳ /tweeter/tweets/tweet[@id=/tweeter/tweets/tweet[@id =  
  ↳ 'T42']/header/answers/answer/@ref]/@author_ref]
```

Chapitre 5

Propriétés des requêtes XPath

5.1 Reformuler les requêtes suivantes en utilisant exclusivement les axes child, descendant, descendant-or-self, following et following-sibling

5.1.1 Requete 1

```
1 //b[parent::a]
```

devient

```
1 //a/b
```

5.1.2 Requete 2

```
1 //a/preceding-sibling::c
```

devient

```
1 //c[following-sibling::a]
```

5.1.3 Requete 3

```
1 //c[preceding::d]
```

devient

```
1 //d/following::c
```

5.1.4 Requete 4

```
1 //b/a/preceding-sibling::c/preceding::d
    devient
1 //d[following::c[following-sibling::a[child::b]]]
```

5.1.5 Requete 5

```
1 /a/b/.../*/.../preceding::d
    devient
1 //d[following::*[child::*]]
```

5.1.6 Requete 6

```
1 //a/ancestor::b/parent::c/child::d/parent::e
```

La requête est syntaxiquement correcte mais ne retourne de résultat pour aucun document XML. Voir explication dans la solution au troisième exercice.

5.2 Reformuler les requêtes `//a/following : :b` et `//a/preceding : :b` en utilisant les axes descendant-or-self, ancestor, following-sibling et preceding-sibling.

5.2.1 Requete 1

```
1 //a/following::b
    devient
1 //a/ancestor-or-self::node()/following-sibling::node()/descendant-or-self::b
```

5.2.2 Requete 2

```
1 //a/preceding::b
    devient
1 //a/ancestor-or-self::node()/preceding-sibling::node()/descendant-or-self::b
```

5.3 Pour chaque requête définie aux points 1 et 2, proposer un document XML pour lequel la réponse à la requête n'est pas vide, sinon expliquer pourquoi un tel document n'existe pas.

5.3.1 Requete 1 - exercice 1

```
1 <r>
2   <a>
3     <b>
4     </b>
5   </a>
6 </r>
```

5.3.2 Requete 2 - exercice 1

```
1 <r>
2   <c>
3   </c>
4   <a>
5   </a>
6 </r>
```

5.3.3 Requete 3 - exercice 1

```
1 <r>
2   <a>
3     <d>
4     </d>
5   </a>
6   <c>
7   </c>
8 </r>
```

5.3.4 Requete 4 - exercice 1

```
1 <r>
2   <a>
3     <d>
4     </d>
5   </a>
6   <c>
7   </c>
```

```

8      <a>
9          <b>
10             </b>
11         </a>
12 </r>

```

5.3.5 Requete 5 - exercice 1

```

1      <r>
2          <d>
3              </d>
4          <a>
5              <b>
6                  </b>
7              </a>
8      </r>

```

5.3.6 Requete 6 - exercice 1

Il n'y a aucun document XML correspondant car la fin de la requête est :

```

1  c/child::d/parent::e

```

Or un noeud de type c ne peut pas avoir d'enfant de type d qui ait un parent e car un noeud n'a qu'un seul parent.

5.3.7 Requete 1 - exercice 2

```

1  <r>
2      <s>
3          <a></a>
4          <b>b1</b>
5      </s>
6      <b>b2</b>
7  </r>

```

5.3.8 Requete 2 - exercice 2

```

1  <r>
2      <b>b2</b>
3      <s>
4          <b>b1</b>
5          <a></a>
6      </s>
7  </r>

```

5.4 Donner un document XML pour lequel la requête `//r[a[1] = a[2]]` n'est pas vide sans que les éléments comparés soient strictement identiques.

```
1  <r>
2    <a>
3      abcdef
4    </a>
5    <a>
6      <b>
7        abc
8      </b>
9      <b>
10     def
11   </b>
12 </a>
13 </r>
```

Les deux noeuds de type `a` sont égaux en XPath car ils partagent la sous-chaine `abc`, pourtant il n'y a pas isomorphisme entre les deux sous-arbres et les noeuds textes ne sont pas identiques.

5.5 Est il vrai que, dans le cadre du langage XPath, si $X = Y$ et $Y = Z$ alors $X = Z$?

Non, en XPath, l'égalité n'est pas transitive.
Voici un contre exemple :

```
1  <root>
2    <x>
3      abc
4    </x>
5    <y>
6      <y1>
7        abc
8      </y1>
9      <y2>
10     def
11   </y2>
12 </y>
13 <z>
14   def
```

```
15     </z>
16 </root>
```

Dans ce document XML, les égalités $x = y$ et $y = z$ sont vraie en XPath mais l'égalité $x = z$ est fausse

Chapitre 6

XQuery : Tweets

Donner les requêtes XQuery correspondants aux expressions suivantes et évaluer ces expressions dans votre document XML contenant des Tweets.

Voir DTD et document XML exemple en annexe.

6.1 Créez une liste de paires tweet-auteur, avec chaque paire contenue dans un element result.

```
1 <results> {  
2   for $auteur in /tweeter/users/user  
3   for $tweet in /tweeter/tweets/tweet  
4   where $auteur/@id = $tweet/@author_ref  
5   return  
6   <result>  
7   { $tweet }  
8   { $auteur }  
9   </result>  
10 }  
11 </results>
```

6.2 Pour chaque utilisateur, listez le nom de l'utilisateur et la date de tous ses tweets, le tout regroupé dans un élément result.

```
1 <results> {  
2   for $auteur in /tweeter/users/user  
3   return  
4   <result>
```

```

5     <nom>
6     { $auteur/user_name/text() }
7     </nom>
8     {
9         for $tweet in /tweeter/tweets/tweet
10        where $auteur/@id = $tweet/@author_ref
11        return $tweet/header/date
12    }
13    </result>
14 }
15 </results>

```

6.3 Listez les utilisateurs qui ont publié un tweet qui a été retwitté au moins deux fois.

```

1 <results> {
2   for $auteur in /tweeter/users/user
3   for $tweet in /tweeter/tweets/tweet
4   where $auteur/@id = $tweet/@author_ref and
5     ↳ count($tweet/header/retweets/retweet) > 1
6   return $auteur
7 }
</results>

```

6.4 Pour chaque tweet, listez son contenu et la date de ses deux premières réponses. Rajoutez un element vide <nonRetwitted/> s'il n'a pas été retwitté.

```

1 <results> {
2   for $tweet in /tweeter/tweets/tweet
3   return
4   <tweet>
5     <contenu>
6     { $tweet/body/text }
7     </contenu>
8     <premieres-reponses>
9     {
10    ↳ /tweeter/tweets/tweet[@id=$tweet/header/answers/answer[1]/@ref]/header/date
11    ↳ }
12    {
13    ↳ /tweeter/tweets/tweet[@id=$tweet/header/answers/answer[2]/@ref]/header/date
14    ↳ }

```

```

11     </premieres-reponses>
12     {
13         if (count($tweet/header/retweets/retweet) = 0) then
14             <nonRetweeted />
15         else()
16     }
17 </tweet>
18 }
19 </results>

```

6.5 Listez les utilisateurs de la plateforme en ordre alphabétique.

```

1 <results> {
2   for $user in tweeter/users/user
3   order by upper-case($user/user_name/text()) ascending
4   return $user
5 }
6 </results>

```

6.6 Listez les tweets contenant l’hashtag “#I;3XML”.

```

1 <results> {
2   for $tweet in /tweeter/tweets/tweet
3   where $tweet/body/hashtag[contains(., "#I&lt;3XML")]
4   return $tweet
5 }
6 </results>

```

6.7 Trouvez le tweet le plus ancien ainsi que le plus recent.

```

1 <results> {
2   let $minDate := min(/tweeter/tweets/tweet/header/date)
3   let $maxDate := max(/tweeter/tweets/tweet/header/date)
4   for $tweet in /tweeter/tweets/tweet
5   where $tweet/header/date = $minDate or $tweet/header/date =
6     ↪ $maxDate
7     return $tweet
8 }
9 </results>

```

6.8 Pour chaque tweet ayant des hashtags, retournez le tweet avec la liste de ses hashtag.

```
1 <results> {  
2   for $tweet in /tweeter/tweets/tweet  
3   where $tweet/body/hashtag  
4   return  
5     <result>  
6       {$tweet}  
7       <hashtags>  
8         {$tweet/body/hashtag}  
9       </hashtags>  
10    </result>  
11  }  
12 </results>
```

6.9 Pour chaque tweet ayant des références utilisateur, retournez le tweet avec la liste des références utilisateur.

```
1 <results> {  
2   for $tweet in /tweeter/tweets/tweet  
3   where $tweet/body/user_ref  
4   return  
5     <result>  
6       {$tweet}  
7       <user_references>  
8         {$tweet/body/user_ref}  
9       </user_references>  
10    </result>  
11  }  
12 </results>
```

6.10 Declarez la fonction local :aReponduAuTweet, qui, étant donné un tweet, retourne tous les utilisateurs qui ont répondu au Tweet.

```
1 declare function local:aReponduAuTweet($tweet)  
2 {  
3   for $user in  
4     ↪ $tweet/parent::tweets/parent::tweeter/child::users/child::user
```

```
4      where $user/@id = $tweet/parent::tweets/child::tweet[@id =  
      ↪ $tweet/header/answers/answer/@ref]/@author_ref  
5      return $user  
6  };
```

Chapitre 7

XQuery : Trains

Donner des expressions XQuery pour les requêtes suivantes.
Voir document XML en annexe.

7.1 Le numéro des trains possédant une voiture-bar.

```
1  <results> {  
2    for $train in /gare/train  
3    where $train/voiture/bar  
4    return  
5    <numero>  
6    {$train/@numero}  
7    </numero>  
8  }  
9  </results>
```

7.2 Le nom des usages ayant au moins une réservation

```
1  <results> {  
2    let $resa := /gare/train/voiture/resa  
3    for $user in /gare/usager  
4    where $user/@id = $resa/@id  
5    return  
6    <usager>  
7    { string($user/@nom) }  
8    </usager>  
9  }  
10 </results>
```

7.3 La reservation avec le plus grand identifiant (dans l'ordre lexicographique).

```
1 <results> {  
2   (for $resa in /gare/train/voiture/resa  
3     order by $resa/@numero descending  
4     return $resa)[1]  
5 }  
6 </results>
```

7.4 Le numéro des trains dont au moins 2 places sont réservées.

```
1 <results> {  
2   for $voiture in /gare/train/voiture  
3   where (count($voiture/resa) >= 2)  
4   return  
5     <train>  
6     { string($voiture/parent::train/@numero) }  
7     </train>  
8 }  
9 </results>
```

7.5 Le nom des personnes ayant réservé exacte- ment deux fois.

```
1 <results> {  
2   for $user in /gare/usager  
3   let $count := count(/gare/train/voiture/resa[@id = $user/@id])  
4   ↪ return  
5     if ($count = 2) then  
6       <usager>  
7       { string($user/@nom) }  
8       </usager>  
9     else()  
10 }  
10 </results>
```

7.6 Les usagers n'ayant effectué aucune réservation.

```
1 <results> {  
2   let $resa := /gare/train/voiture/resa  
3   for $user in /gare/usager
```

```
4     where not($user/@id = $resa/@id)
5     return $user
6 }
7 </results>
```


Annexe A

DTD - Films

```
1  <!DOCTYPE FILMS [  
2  <!ELEMENT FILMS (FILM+, ARTISTE+)>  
3  <!ELEMENT FILM (TITRE, ANNEE, GENRE, PAYS, MES, ROLES,  
   ↪ RESUME?)>  
4  <!ELEMENT TITRE (#PCDATA)>  
5  <!ELEMENT ANNEE (#PCDATA)>  
6  <!ELEMENT GENRE (#PCDATA)>  
7  <!ELEMENT PAYS (#PCDATA)>  
8  <!ELEMENT MES (#PCDATA)>  
9  <!ATTLIST MES idref CDATA #REQUIRED>  
10 <!ELEMENT ROLES (ROLE*)>  
11 <!ELEMENT ROLE (#PCDATA)>  
12 <!ATTLIST ROLE idref CDATA #REQUIRED>  
13 <!ELEMENT RESUME (#PCDATA)>  
14 <!ELEMENT ARTISTE (ACTNOM, ACTPNOM, ANNEENAISS)>  
15 <!ATTLIST ARTISTE id CDATA #REQUIRED>  
16 <!ELEMENT ACTNOM (#PCDATA)>  
17 <!ELEMENT ACTPNOM (#PCDATA)>  
18 <!ELEMENT ANNEENAISS (#PCDATA)>  
19 ]>
```

Annexe B

XML - Recettes

```
1  <recettes>
2  <recette nomCourt="Chiffonnade" nom="Chiffonnade de jambon et
   ↳ d'asperges à la Flamande" type="salee">
3    <materiel>
4      <ingredient quantite="8">asperge</ingredient>
5      <ingredient quantite="150g">jambon fume</ingredient>
6      <ingredient quantite="2">oeuf</ingredient>
7      <ingredient quantite="6 cl">huile d'olive</ingredient>
8      <ingredient quantite="1 c. à soupe">persil
   ↳ hache</ingredient>
9      <ingredient>poivre</ingredient>
10     <ingredient>sel</ingredient>
11     <ingredient>noix de muscade</ingredient> </materiel>
   ↳ <methode> A l'aide du hache-legumes, raper les asperges
   ↳ en lanières d'environ 1,5 mm d'épaisseur et les cuire
   ↳ dans l'eau salee. Couper également les tranches de
   ↳ jambon en longues lanières et les melanger aux asperges
   ↳ cuites et tiedies.
12   </methode>
13 </recette>
14 <recette nomCourt="Pain à l'huile" nom="Presse d'olive sur lit
   ↳ de ble" type="salee">
15   <materiel>
16     <ingredient quantite="1 baguette">pain</ingredient>
17     <ingredient quantite="3 c. à soupe">huile
   ↳ d'olive</ingredient>
18   </materiel>
```

```
19      <methode> A l'aide d'un couteau effile, trancher la baguette
      ↪ sur toute sa longueur. Badigeonner delicatement chaque
      ↪ tranche avec l'huile, et servir immédiatement. Attention,
      ↪ ce plat constitue un repas complet, tout dessert est
      ↪ inutile.
20      </methode>
21  </recette>
22 </recettes>
```

Annexe C

XML - Trains

```
1 <gare>
2   <train numero="t5560" type="TGV">
3     <voiture numero="v1">
4       <resa numero="r17" id="u55"/>
5       <resa numero="r18" id="u52"/>
6     </voiture>
7     <voiture numero="v2"/>
8     <voiture numero="v3"/>
9     <voiture numero="v4">
10      <bar service="froid uniquement"/>
11    </voiture>
12  </train>
13  <train numero="t6731">
14    <voiture numero="v1"/> 2
15    <voiture numero="v2">
16      <resa numero="r15" id="u55"/>
17    </voiture>
18  </train>
19  <usager id="u55" nom="Jean" prenom="Dufour"/>
20  <usager id="u52" nom="Brigitte" prenom="Lefebvre"/>
21  <usager id="u56" nom="Patrick" prenom="Subiran"/>
22 </gare>
```

Annexe D

DTD - Tweets

```
1  <!ELEMENT tweeter (users, tweets) >
2  <!ELEMENT users (user*)>
3  <!ELEMENT user (user_name, first_name, last_name, profile)>
4  <!ATTLIST user id ID #REQUIRED >
5  <!ELEMENT user_name (#PCDATA) >
6  <!ELEMENT first_name (#PCDATA) >
7  <!ELEMENT last_name (#PCDATA) >
8  <!ELEMENT profile (#PCDATA) >
9
10 <!ELEMENT tweets (tweet*)>
11 <!ELEMENT tweet (header, body)>
12 <!ATTLIST tweet id ID #REQUIRED >
13 <!ATTLIST tweet author_ref IDREF #REQUIRED >
14 <!ELEMENT header (date, timezone, location, language, retweets,
    ↪  answers, operating_system, images*, videos*) >
15 <!ELEMENT date (#PCDATA) >
16 <!ELEMENT timezone (standard, offset) >
17 <!ELEMENT standard (#PCDATA) >
18 <!ELEMENT offset (#PCDATA) >
19 <!ELEMENT location (latitude, longitude, city, country) >
20 <!ELEMENT latitude (#PCDATA)>
21 <!ELEMENT longitude (#PCDATA)>
22 <!ELEMENT city (#PCDATA)>
23 <!ELEMENT country (#PCDATA)>
24 <!ELEMENT language (#PCDATA)>
25 <!ELEMENT retweets (retweet*)>
26 <!ELEMENT retweet EMPTY>
27 <!ATTLIST retweet ref IDREF #REQUIRED >
28 <!ELEMENT answers (answer*)>
29 <!ELEMENT answer EMPTY >
30 <!ELEMENT operating_system (#PCDATA)>
```

```

31 <!ELEMENT images (image*)>
32 <!ELEMENT image (#PCDATA)>
33 <!ATTLIST image id ID #REQUIRED >
34 <!ELEMENT videos (video*)>
35 <!ELEMENT video (#PCDATA)>
36 <!ATTLIST video id ID #REQUIRED >
37 <!ATTLIST answer ref IDREF #REQUIRED >
38 <!ELEMENT body (text|hashtag|image_ref|user_ref|video_ref)+ >
39 <!ELEMENT text (#PCDATA)>
40 <!ELEMENT hashtag (#PCDATA)>
41 <!ELEMENT image_ref EMPTY>
42 <!ATTLIST image_ref ref IDREF #REQUIRED >
43 <!ELEMENT video_ref EMPTY>
44 <!ATTLIST video_ref ref IDREF #REQUIRED >
45 <!ELEMENT user_ref (#PCDATA)>
46 <!ATTLIST user_ref ref IDREF #REQUIRED >

```

Annexe E

XML - Tweets

```
1  <tweeter>
2  <users>
3    <user id="U41">
4      <user_name>
5        basil_dalie
6      </user_name>
7      <first_name>
8        Basil
9      </first_name>
10     <last_name>
11       Dalie
12     </last_name>
13     <profile>
14       https://twitter.com/ basil_dalie
15     </profile>
16   </user>
17   <user id="U02">
18     <user_name>
19       Cicero
20     </user_name>
21     <first_name>
22       Ci
23     </first_name>
24     <last_name>
25       Cero
26     </last_name>
27     <profile>
28       https://twitter.com/Cicero
29     </profile>
30   </user>
31   <user id="U43">
```

```

32     <user_name>
33         alex_not
34     </user_name>
35     <first_name>
36         Alex
37     </first_name>
38     <last_name>
39         Not
40     </last_name>
41     <profile>
42         https://twitter.com/alex_n
43     </profile>
44 </user>
45 </users>
46 <tweets>
47     <tweet id="T42" author_ref="U41">
48         <header>
49             <date>
50                 1758312000
51             </date>
52             <timezone>
53                 <standard>
54                     UTC
55                 </standard>
56                 <offset>
57                     1
58                 </offset>
59             </timezone>
60             <location>
61                 <latitude>
62                     3.876716
63                 </latitude>
64                 <longitude>
65                     43.610769
66                 </longitude>
67                 <city>
68                     Montpellier
69                 </city>
70                 <country>
71                     France
72                 </country>
73             </location>
74             <language>
75                 Latin
76             </language>
77             <retweets>

```



```

78         <retweet ref="T04" />
79         <retweet ref="T05" />
80         <retweet ref="T03" />
81     </retweets>
82     <answers>
83         <answer ref="T02" />
84     </answers>
85     <operating_system>
86         Linux x86_64
87     </operating_system>
88     <images>
89         <image id="I01">
90             ↪ https://images2.minutemediacdn.com/image/upload/c_crop,h_1193,w_2121,x_0,y_0
91         </image>
92         <image id="I02">
93             https://website.com/image.jpg
94         </image>
95     </images>
96     <videos>
97         <video id="V01">
98             https://website.com/video.avi
99         </video>
100    </videos>
101 </header>
102 <body>
103     <text>
104         Lorem ipsum dolor sit amet, consectetur adipiscing
105         ↪ elit, sed do eiusmod tempor incididunt ut labore et
106         ↪ dolore magna aliqua. Ut enim ad minim veniam, quis
107         ↪ nostrud exercitation ullamco laboris nisi ut
108         ↪ aliquip ex ea commodo consequat. Duis aute irure
109         ↪ dolor in reprehenderit in voluptate velit esse
110         ↪ cillum dolore eu fugiat nulla pariatur. Excepteur
111         ↪ sint occaecat cupidatat non proident, sunt in culpa
112         ↪ qui officia deserunt
113     </text>
114     <text>
115         anim id est laborum
116     </text>
117     <image_ref ref="I01" />
118     <user_ref ref="U02">
119         @Cicero
120     </user_ref>
121     <text>

```

```

114      Sed ut perspiciatis unde omnis iste natus error sit
      ↪ voluptatem accusantium doloremque laudantium, totam
      ↪ rem aperiam, eaque ipsa quae ab illo inventore
      ↪ veritatis et quasi architecto beatae vitae dicta
      ↪ sunt explicabo. Nemo enim ipsam voluptatem quia
      ↪ voluptas sit aspernatur aut odit aut fugit, sed
      ↪ quia consequuntur magni dolores eos qui ratione
      ↪ voluptatem sequi nesciunt. Neque porro quisquam
      ↪ est, qui dolorem ipsum quia dolor sit amet,
      ↪ consectetur, adipisci velit, sed quia non numquam
      ↪ eius modi tempora incidunt ut labore et dolore
      ↪ magnam aliquam quaerat voluptatem. Ut enim ad
      ↪ minima veniam, quis nostrum exercitationem ullam
      ↪ corporis suscipit laboriosam, nisi ut aliquid ex ea
      ↪ commodi consequatur? Quis autem vel eum iure
      ↪ reprehenderit qui in ea voluptate velit esse quam
      ↪ nihil molestiae consequatur, vel illum qui dolorem
      ↪ eum fugiat quo voluptas nulla pariatur?
115      </text>
116      <video_ref ref="V01" />
117    </body>
118  </tweet>
119  <tweet id="T02" author_ref="U02" >
120    <header>
121      <date>
122        1758311999
123      </date>
124      <timezone>
125        <standard>
126          UTC
127        </standard>
128        <offset>
129          1
130        </offset>
131      </timezone>
132      <location>
133        <latitude>
134          3.876716
135        </latitude>
136        <longitude>
137          43.610769
138        </longitude>
139        <city>
140          Montpellier
141        </city>
142        <country>

```

```

143         France
144     </country>
145 </location>
146 <language>
147     Latin
148 </language>
149 <retweets>
150 </retweets>
151 <answers>
152     <answer ref="T42" />
153     <answer ref="T43" />
154     <answer ref="T44" />
155 </answers>
156 <operating_system>
157     Linux x86_64
158 </operating_system>
159 </header>
160 <body>
161     <hashtag>
162         #mollit
163     </hashtag>
164     <text>
165         Lorem ipsum dolor sit amet, consectetur adipiscing
166         ↪ elit, sed do eiusmod tempor incididunt ut labore et
167         ↪ dolore magna aliqua. Ut enim ad minim veniam, quis
168         ↪ nostrud exercitation ullamco laboris nisi ut
169         ↪ aliquip ex ea commodo consequat. Duis aute irure
170         ↪ dolor in reprehenderit in voluptate velit esse
171         ↪ cillum dolore eu fugiat nulla pariatur. Excepteur
172         ↪ sint occaecat cupidatat non proident, sunt in culpa
173         ↪ qui officia deserunt
174     </text>
175     <hashtag>
176         #I&lt;3XML
177     </hashtag>
178     <text>
179         anim id est laborum
180     </text>
181     <video_ref ref="I01" />
182 </body>
183 </tweet>
184 <tweet id="T03" author_ref="U41" >
185     <header>
186         <date>
187             1758312001
188         </date>

```

```

181     <timezone>
182         <standard>
183             UTC
184         </standard>
185         <offset>
186             1
187         </offset>
188     </timezone>
189     <location>
190         <latitude>
191             3.876716
192         </latitude>
193         <longitude>
194             43.610769
195         </longitude>
196         <city>
197             Montpellier
198         </city>
199         <country>
200             France
201         </country>
202     </location>
203     <language>
204         Latin
205     </language>
206     <retweets />
207     <answers />
208     <operating_system>
209         Linux x86_64
210     </operating_system>
211 </header>
212 <body>
213     <text>
214         Lorem ipsum dolor sit amet, consectetur adipiscing
        ↪ elit, sed do eiusmod tempor incididunt ut labore et
        ↪ dolore magna aliqua. Ut enim ad minim veniam, quis
        ↪ nostrud exercitation ullamco laboris nisi ut
        ↪ aliquip ex ea commodo consequat. Duis aute irure
        ↪ dolor in reprehenderit in voluptate velit esse
        ↪ cillum dolore eu fugiat nulla pariatur. Excepteur
        ↪ sint occaecat cupidatat non proident, sunt in culpa
        ↪ qui officia deserunt
215     </text>
216     <hashtag>
217         nofilter
218     </hashtag>

```

```

219         <hashtag>
220             xml
221         </hashtag>
222     </body>
223 </tweet>
224 <tweet id="T43" author_ref="U02" >
225     <header>
226         <date>
227             1758312002
228         </date>
229         <timezone>
230             <standard>
231                 UTC
232             </standard>
233             <offset>
234                 1
235             </offset>
236         </timezone>
237         <location>
238             <latitude>
239                 3.876716
240             </latitude>
241             <longitude>
242                 43.610769
243             </longitude>
244             <city>
245                 Montpellier
246             </city>
247             <country>
248                 France
249             </country>
250         </location>
251         <language>
252             Latin
253         </language>
254         <retweets />
255         <answers />
256         <operating_system>
257             Linux x86_64
258         </operating_system>
259     </header>
260     <body>
261         <text>

```

```

262         Lorem ipsum dolor sit amet, consectetur adipiscing
           ↳ elit, sed do eiusmod tempor incididunt ut labore et
           ↳ dolore magna aliqua. Ut enim ad minim veniam, quis
           ↳ nostrud exercitation ullamco laboris nisi ut
           ↳ aliquip ex ea commodo consequat. Duis aute irure
           ↳ dolor in reprehenderit in voluptate velit esse
           ↳ cillum dolore eu fugiat nulla pariatur. Excepteur
           ↳ sint occaecat cupidatat non proident, sunt in culpa
           ↳ qui officia deserunt
263     </text>
264 </body>
265 </tweet>
266 <tweet id="T44" author_ref="U02" >
267     <header>
268         <date>
269             1758312003
270         </date>
271         <timezone>
272             <standard>
273                 UTC
274             </standard>
275             <offset>
276                 1
277             </offset>
278         </timezone>
279         <location>
280             <latitude>
281                 3.876716
282             </latitude>
283             <longitude>
284                 43.610769
285             </longitude>
286             <city>
287                 Montpellier
288             </city>
289             <country>
290                 France
291             </country>
292         </location>
293         <language>
294             Latin
295         </language>
296         <retweets />
297         <answers />
298         <operating_system>
299             Linux x86_64

```

```

300     </operating_system>
301 </header>
302 <body>
303     <text>
304         Lorem ipsum dolor sit amet, consectetur adipiscing
            ↳ elit, sed do eiusmod tempor incididunt ut labore et
            ↳ dolore magna aliqua. Ut enim ad minim veniam, quis
            ↳ nostrud exercitation ullamco laboris nisi ut
            ↳ aliquip ex ea commodo consequat. Duis aute irure
            ↳ dolor in reprehenderit in voluptate velit esse
            ↳ cillum dolore eu fugiat nulla pariatur. Excepteur
            ↳ sint occaecat cupidatat non proident, sunt in culpa
            ↳ qui officia deserunt
305     </text>
306 </body>
307 </tweet>
308 <tweet id="T04" author_ref="U02" >
309     <header>
310         <date>
311             1758312004
312         </date>
313         <timezone>
314             <standard>
315                 UTC
316             </standard>
317             <offset>
318                 1
319             </offset>
320         </timezone>
321         <location>
322             <latitude>
323                 3.876716
324             </latitude>
325             <longitude>
326                 43.610769
327             </longitude>
328             <city>
329                 Montpellier
330             </city>
331             <country>
332                 France
333             </country>
334         </location>
335         <language>
336             Latin
337         </language>

```

```

338     <retweets />
339     <answers />
340     <operating_system>
341         Linux x86_64
342     </operating_system>
343 </header>
344 <body>
345     <text>
346         Lorem ipsum dolor sit amet, consectetur adipiscing
            ↳ elit, sed do eiusmod tempor incididunt ut labore et
            ↳ dolore magna aliqua. Ut enim ad minim veniam, quis
            ↳ nostrud exercitation ullamco laboris nisi ut
            ↳ aliquip ex ea commodo consequat. Duis aute irure
            ↳ dolor in reprehenderit in voluptate velit esse
            ↳ cillum dolore eu fugiat nulla pariatur. Excepteur
            ↳ sint occaecat cupidatat non proident, sunt in culpa
            ↳ qui officia deserunt
347     </text>
348 </body>
349 </tweet>
350     <tweet id="T05" author_ref="U02" >
351 <header>
352     <date>
353         1758312005
354     </date>
355     <timezone>
356         <standard>
357             UTC
358         </standard>
359         <offset>
360             1
361         </offset>
362     </timezone>
363     <location>
364         <latitude>
365             3.876716
366         </latitude>
367         <longitude>
368             43.610769
369         </longitude>
370         <city>
371             Montpellier
372         </city>
373         <country>
374             France
375         </country>

```



```

376         </location>
377         <language>
378             Latin
379         </language>
380         <retweets />
381         <answers />
382         <operating_system>
383             Linux x86_64
384         </operating_system>
385     </header>
386     <body>
387         <text>
388             Lorem ipsum dolor sit amet, consectetur adipiscing
389                 ↳ elit, sed do eiusmod tempor incididunt ut labore et
390                 ↳ dolore magna aliqua. Ut enim ad minim veniam, quis
391                 ↳ nostrud exercitation ullamco laboris nisi ut
392                 ↳ aliquip ex ea commodo consequat. Duis aute irure
393                 ↳ dolor in reprehenderit in voluptate velit esse
394                 ↳ cillum dolore eu fugiat nulla pariatur. Excepteur
395                 ↳ sint occaecat cupidatat non proident, sunt in culpa
396                 ↳ qui officia deserunt
397         </text>
398     </body>
399 </tweet>
400 </tweets>
401 </tweeter>

```