

Données du Web - TD 2 : XML et DTDs

TD en binômes. Rendu facultatif possible sur Moodle, le 24/09.

1) Instance de données

- Représentez le Tweet suivant dans un document XML valide par rapport à la DTD que vous avez proposée pour le TD précédent.



2) Synthèse des modèles

- Unifiez votre schéma avec celui d'un autre binôme, en proposant un seul modèle pour les deux groupes.
- Illustrez (au plus) trois différences entre les deux modèles, ainsi que la solution retenue dans la synthèse des modèles.

3) DTD : cas particuliers

Pour chaque DTD, donner un arbre XML valide. Sinon, expliquer pourquoi un tel arbre n'existe pas.

DTD 1

```
<!DOCTYPE B [
  <!ELEMENT B (A,C) >
  <!ELEMENT C (D) >
  <!ELEMENT A (D) >
  <!ELEMENT C #PCDATA >
  <!ELEMENT D #PCDATA > ]>
```

DTD 2

```
<!DOCTYPE EMPTY [
  <!ELEMENT EMPTY EMPTY > ]>
```

DTD 3

```
<!DOCTYPE A [
  <!ELEMENT A (B) >
  <!ELEMENT B (A) > ]>
```

DTD 4.

```
<!DOCTYPE C [
  <!ELEMENT C (C*) > ]>
```

DTD 5.

```
<!DOCTYPE C [
  <!ELEMENT C (C,EMPTY)* >
  <!ELEMENT EMPTY EMPTY > ]>
```

DTD 6.

```
<!DOCTYPE C [
  <!ELEMENT C (B*, C, C, C, C) > ]>
```

5) Expressions régulières et déterminisme

Déterminer si les expressions régulières suivantes sont déterministes.

$$r_1 = (a^*(b|a))(b, (a|c)) \quad r_2 = (a, (a|b)^+)(c, (a|b)) \quad r_3 = (a|(a, a, a))^*, (a|(d, d, b))^*, e$$

6) La simplification des expressions régulières

Les expressions régulières suivantes r_1, r_2, r_3 sont équivalentes à $r = (a|b)^*$.

$$\begin{aligned} r_1 &= (a^*, b^*)^* \\ r_2 &= (a^?, (a|b)^*) \\ r_3 &= ((a|a)^+ | b)^* \end{aligned}$$

Dans une DTD, on peut donc remplacer toute instance de r_1, r_2, r_3 par r .

Question : à l'aide du graphe associé à une expression régulière, proposer un algorithme permettant de vérifier si une expression utilisant les symboles $\{a_1, a_2, \dots, a_n\}$ est équivalente à $(a_1|a_2|\dots|a_n)^*$.