

INFO-H-509 : Technologies XML

TP 1 - Introduction à XML et XPath

Professeur : Stijn Vansummeren
Assistant : Michaël Waumans
<http://cs.ulb.ac.be/public/teaching/infoh509>
2015-2016

XML Well-formedness

Les exercices font références au document suivant, aussi disponible en téléchargement sur la page des TPs. Ce document vise à décrire un catalogue de cours :

```
<?xml version="1.0" encoding="utf-8" ?>
<catalogue>
  <course mnemonic="INFO-H-509">
    <title>Technologies XML
    <professor email="stijn.vansummeren@ulb.ac.be">Stijn Vansummeren</prof>
    <assistant email="fpicalau@ulb.ac.be">François Picalausa</assistant>
    <weight>3 <unit type="ECTS" /></weight>
  </course>
</catalogue>
<catalogue>
  <course mnemonic="INFO-H-302" <!-- Ce cours ne sera plus donné l'année
prochaine --> >
    <title>Analyse & conception par objet</title>
    <professor email="ezimanyi@ulb.ac.be">Esteban Zimányi
    <assistant email="fservais@ulb.ac.be">Frédéric Servais</assistant>
    < assistant email="sboucher@ulb.ac.be">Serge Boucher</assistant>
    <weight>3 <unit type="ECTS"></weight>
  </course>
</catalogue>
```

Exercice 1.1

Corriger le document fournit pour qu'il soit *Well-formed*.

Exercice 1.2

Compléter ce document pour inclure le cours de *Bases de données*, INFO-H-303. Ce cours est enseigné par le prof. Zimányi, et Frédéric Servais en est l'assistant. Il s'agit d'un cours de 6 ECTS.

Exercice 1.3

Compléter ce document en ajoutant plusieurs éléments **person**, de sorte à reprendre les informations suivantes. Utilisez judicieusement éléments et attributs.

Nom	Laboratoire	Email	Domaines de recherche
Stijn Vansummeren	WIT	stijn.vansummeren@ulb.ac.be	Systèmes d'information Bases de données scientifiques Web sémantique
Esteban Zimányi	WIT	ezimanyi@ulb.ac.be	Systèmes d'information Entrepôts de données Web sémantique
François Picalausa	WIT	fpicalau@ulb.ac.be	Web sémantique

Unicode

Les pages suivantes sont à utiliser pour réaliser les exercices Unicode :

- <http://www.unicode.org/charts/index.html>
- <http://www.fileformat.info/info/unicode/>

Exercice 1.4

Donner le caractère associé au *code point* $265C_{16}$ (9820 en décimal). Donner le nom et le glyphe.

Exercice supplémentaire : donner l'encodage UTF-8 de ce caractère.

Exercice 1.5

Donner les code points encodés par la séquence UTF-8 00110100 00110010 11100010 10000000 10100110.

Requêtes XPath

Télécharger le jeu de données “orders.xml” sur la page des TPs.

Exercice 1.6

Exécuter et expliquer les requêtes suivantes :

1. `//Customer[@CustomerID="HUNGC"]//Country`
2. `/Root/Orders//ShipCountry | /Root/Customers//Country`
3. `//ShipCity[following-sibling::ShipRegion eq "OR"]`
4. `//ShipCity[following-sibling::ShipRegion is "OR"]`
5. `/Root/Customers/Customer/Phone[1]`
6. `(/Root/Customers/Customer/Phone)[1]`

Exercice 1.7

Réaliser les requêtes suivantes :

1. Les clients qui ont le titre de Marketing Manager.
2. Les éléments HTML du document.
3. Les clients dont le nom *contient* Yoshi.
4. Les frais de transports les plus élevés, les moins élevés et moyens. (Élément Freight)
5. Les commandes qui n'ont pas encore été envoyées (attribut shippedDate absent pour shipInfo)
6. La première commande (dans l'ordre du document), pour chaque client.
7. Les commandes réalisées par GREAL en Avril 1998.
8. Les commandes réalisées par un client basé en californie.
9. La liste des pays de livraison ou, pour l'Amérique, des états.
10. La requête qui teste s'il existe une commande livrée par le livreur 3.