

INFO-H-509 : Technologies XML

TP 1 - An introduction to XML and XPath

Professor : Stijn Vansummeren
Teaching Assistant : Michaël Waumans
<http://cs.ulb.ac.be/public/teaching/infoh509>

2013-2014

XML Well-formedness

The following exercises refer to the following XML document that describes a catalog of courses (it is also available on the web page).

```
<?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 object</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>
```

Exercise 1.1

Is the document well-formed? If not, make the necessary corrections so that it becomes well-formed.

Exercise 1.2

Complete the document to include a new course *Bases de données*, INFO-H-303. This course is taught by Prof. Zimanyi. Frederic Servais is the teaching assistant. This is a course of 6 ECTS.

Exercise 1.3

Complete this document with several elements `person`, in order to include the following information. You have to use wisely elements and attributes.

Name	Laboratory	Email	Fields of Research
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

The following web pages must be used to solve the following exercises on Unicode :

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

Exercise 1.4

Give the character associated to the *code point* 265C₁₆ (9820 in decimal). Give the name and the glyph.

Additional exercise : give the UTF-8 encoding of this character.

Exercise 1.5

Give code points encoded by the following UTF-8 sequence : 00110100 00110010 11100010 10000000 10100110.

XPath Query

Download the dataset “orders.xml” on the web page.

Exercise 1.6

Execute and explain the following XPath queries :

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]`

Exercise 1.7

Write XPath expressions for the following queries

1. Customers who have the title Marketing Manager.
2. The HTML elements of this document.
3. Customers whose name *contains* Yoshi.
4. The highest, lowest and average freight costs.
5. Orders that have not yet been sent (attribute shippedDate is missing for shipInfo)
6. The first order (in document order), for each customer.
7. Orders made by Greal in April 1998.
8. Orders made by a customer based in California.
9. The countries of delivery (or states for America).
10. The query that tests whether there is a command supplied by delivery man 3.