Tutorial – 1
XML and Semantic Web Technologies – SoSe 2012
Instructor: Prof. Dr. Dr. Lars Schmidt-Thieme
Tutor: Umer Khan 27-04-2012

Solution

Q1. (a) Using UTF-8, decode the following byte sequence into its Unicode code point.
To find code point, bytes are converted to binary.

```
0xE2        0x98        0x83
11100010   10011000    10000011
1110xxxx   10xxxxyy    10yyyyyy
```
Leading bits used to identify start of a 3-bytes sequence and continuation bits are removed

```
xxxx  xxxxyy  yyyyyy
0010  011000  000011
```

Convert them back to hex.

```
xxxxxxxx | yyyyyyyy
00100110 | 00000011
  0x26   | 0x03
```

Unicode: U+ 2603 [Snowman symbol]

(b). Briefly differentiate between Coded Character Set and Encoding Scheme.

Coded character set is a collection of coded characters to represent textual information.

Character encoding scheme is the interpretation of coded character into bytes.

(c) Using example URIs, differentiate between hierarchical, opaque, relative and absolute URIs'.

Absolute URI:

An absolute URL typically takes the following form:

protocol://hostname/other_information
[scheme:]scheme-specific-part[#fragment]

The `protocol` is usually `http://`, but can also be `ftp://` or `file://`.

The `hostname` is the name of the computer / server which hosts the resource. For example, the hostname of University of Hildesheim central web server is `www.uni-hildesheim.de`

The `other_information` includes directory and file information.

You must use absolute URLs when referring to links on different servers.

Relative URI:

Are used only to reference links on the same server as the page that contains them.

The Wonderful World of Foobar!
The Wonderful World of Foobar!

Hierarchical and Opaque URI:

An *opaque* URI is an absolute URI whose scheme-specific part does not begin with a slash character (`'/'`).

```
mailto:java-net@java.sun.com
news:comp.lang.java
```

A *hierarchical* URI is either an absolute URI whose scheme-specific part begins with a slash character, or a relative URI, that is, a URI that does not specify a scheme.

```
http://java.sun.com/j2se/1.3/
docs/guide/collections/designfaq.html#28
```

Q2. (a) Install Oxygen xml editor and play around with its different features.

(b) Is the following xml document 'Well-formed'? If not, point out reasons.

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <course ref="ISMLL">
3       <title>XML & Semantic Web Technologies</title>
4       <people responsible="Prof. Lars">
5              <sub>André Busche</sub>
6       </people role="HoO">
7       <students restrict="grade(Machine Learning) < 4.0)">
8              <student1 formerLecture="Machine Learning" formerLecture="Java1">
9  your name here ... :-> </student1>
10             <student2 formerLecture="Java1 > Machine Learning"/>
```

```
11        </student>
12 </course>
```