

XML Technologies: Exercise 1

Introduction to XML, DTDs

This is a mandatory exercise and the result will be part of your final mark. The solution must be uploaded to OLAT by **March 4th at 15:59**. Late submissions will not be accepted.

Submit the following files in a zipped archive:

- wellformed.xml
- validation.txt
- singlemalt.dtd
- valid.xml
- invalid.xml

Name the archive [lastname]_[firstname]_1.zip (for example *mueller_mathias_1.zip*). Some parts of your submission may be automatically evaluated, so make sure to name your files *precisely* as prescribed, otherwise you might not get any points.

1 Introducing XML and xmllint

1.1 xmllint (0 Points)

`xmllint` is a program that parses XML files and, among other things, can perform *validation*. It is installed by default on most UNIX-based operating systems, and by extension, on OSX, too.

If you are using one of these operating systems, open your console and type:

```
xmllint --version
```

For Windows, there are some instructions¹ for how to install it. However, we cannot offer any support for these as part of this course.

¹<https://www.zlatkovic.com/libxml.en.html#install>

1.2 Well-Formedness (2 Points)

Not all structured data is XML: The property of being “correct” XML is called well-formedness. The document `malformed.xml` is not well-formed. But what precisely is wrong?

```
<?xml version="1.0" encoding="utf-8">
  <company>
    <person role="master distiller">
      <salary >10000</salary>
      <weight unit="kg">90</weight>
    </person>
    <person role="CTO & Business Analyst" role="CTO">
      <salary >9500 </ salary >
      <weight unit=kg">100</weight >
  </company/>
```

Find all mistakes in the attempted XML document above and correct them. Save a well-formed version of it as `wellformed.xml`, explaining your corrections using `<!-- comments -->`. If all your objections are amended, the resulting document should indeed be well-formed XML.

You may use `xmllint` to check your document for well-formedness. In order to do so, open your console, navigate to your directory and type:

```
xmllint --noout wellformed.xml
```

If `xmllint` doesn't complain, then the file is well-formed.

2 Validation and DTD

2.1 Validation and Valid Documents (1 Point)

XML is not a markup language in its strict sense, but a **meta-markup language**. This is because XML does define a syntax to structure data, but does not define a vocabulary. A vocabulary is a collection of key words in the data that map to specific actions in applications. Likewise, an application might expect data items to be stored in a certain order, while XML allows the children of an element in any order.

With all this in mind,

- What is validation and what purpose does it serve?
- Under what conditions can an XML document be said to be valid?
- Can XML documents be well-formed, but not valid?
- Can XML documents be valid, but not well-formed?

Submit your answers in the text file `validation.txt`.

2.2 Document Type Definitions (DTDs) (2 Points)

Below is an example of a short, well-formed XML document. Remember, well-formedness has no say in whether the content makes sense.

```
<?xml version="1.0" encoding="utf-8"?>
<single_malt>
  <whisky name="macallan" origin="Scotland" age="12"/>
  <whiskey name="jameson" origin="Ireland" age="4"/>
  <whisky name="lagavulin" origin="Scotland"/>
  <bourbon name="four_roses" origin="USA" age="5" />
</single_malt>
```

Single Malt must always be made from malted barley, and Bourbon is made from corn. Write a DTD `singlemalt.dtd` where the following rules are laid down:

- A `single_malt` element may only contain `whisky` or `whiskey` elements.
- Every `whisky` and `whiskey` element must have an `age` attribute.
- Every `whisky` and `whiskey` element may have an `origin` and `name` attribute.

Thus, while your DTD should *invalidate* the XML above, the following document should be valid in regard to the DTD you have written:

```
<?xml version="1.0" encoding="utf-8"?>
<single_malt>
  <whisky name="macallan" origin="Scotland" age="12"/>
  <whiskey name="jameson" origin="Ireland" age="4"/>
  <whisky name="lagavulin" origin="Scotland" age="16"/>
</single_malt>
```

Please may test your DTD by using `xmllint` as follows (or resort to some other method of validation):

```
xmllint --noout --dtdvalid singlemalt.dtd valid.xml
```

Also, to facilitate evaluation of your submission, also include the files `valid.xml` and `invalid.xml` in your submission.