Workshop Week 3 - COMP20008

XML

1. What is the difference between XML and HTML ? When would it be more appropriate to use XML instead of HTML ? When would it be more appropriate to use HTML instead of XML ?

2. Type in (or copy and paste) the following XML data using a text editor.

```
<?xml version="1.0" encoding="utf-8"?>

<queen title="Queen Elizabeth II" marriedTo="Philip, Duke of Edinburgh">
    <prince title="Charles, Prince of Wales" marriedTo="Lady Diana Spencer">
    <prince title="Prince William of Wales" />
    <prince title="Prince Henry of Wales" />
    </Prince>
    <princess title="Anne, Princess Royal" />
    <prince title="Andrew, Duke of York" />
    <prince title="Edward, Earl of Wessex" >
</queen>
```

Save the file and name it royal.xml. Load the file in the browser. Notice that the browser would display some errors. In fact, there are actually two syntax errors in the documents. The XML file is not well-formed. Find those errors and fix them. Save the file as *royal2.xml* (put in the same directory as *royal.xml*) and load it in the browser to check it works.

3. Examine the corrected file *royal2.xml* and answer the following questions:

   - Draw the XML tree that corresponds to this XML fragment
   - How many XML elements are there in the whole XML tree? What attributes belong to the first child of the root element? What are their values?
   - Why do you think *title* is an attribute and not an element? Under what circumstances would it be worthwhile to make it an element?
   - Why is *prince* an element and not an attribute? Are there circumstances where it would be worthwhile to make it an attribute?

4. Update *royal2.xml* so that the *prince* element uses the namespace `http://princely.gov.uk`, the *queen* element uses the namespace `http://queenly.gov.uk`, the *princess* element uses `princessly.gov.uk` and all the attributes use the namespace `http://info.gov.uk` Save the file as *royal2-ns.xml* and view it again in your browser to check its correct.

5. Now load up the Jupyter notebook file *XMLandJSON-exercise.ipynb* and do exercises 5a) and 5b). Make sure the *royal2.xml* is in the same directory as the notebook file.

JSON

Consider the following description of a book using XML

```
<?xml version="1.0" encoding="utf-8"?>
<book id="book001">
  <author>Salinger, J. D.</author>
  <title>The Catcher in the Rye</title>
  <price>44.95</price>
  <language>English</language>
  <publish_date>1951-07-16</publish_date>
  <publisher>Little, Brown and Company</publisher>
  <isbn>0-316-76953-3</isbn>
  <description>A story about a few important days in the life of Holden Caulfield</description>
</book>
```

6. Represent the XML file as JSON. Create the new file, give it the name *book.json*, and save as 'Text'. Begin with the following text and expand from there:

   ```
   {
     "id": "book001",
     "author": "Salinger, J. D."
   }
   ```

   Validate your JSON solution against JSONLint.

7. In your JSON solution, add Spanish and German as two extra languages represented as an array. Save this file as *book2.json*. Validate it on JSONLint.

   Now modify the publish_date parameter. Make this an array of two objects that have properties of edition (first, second) and date (1951-07-16,1979-01-01) respectively. Save this file as *book3.json*. Validate it on JSONLint.

More practice with pandas

Only attempt this section during the workshop if there is time remaining. Otherwise, it is an exercise for you to work through in your own time outside of class.

8. This question continues the scenario from the workshop in week 2. The final task is now to study the affect of population size on CO2 emission.

   - Create a new DataFrame object using pd.read_csv('population.csv',index_col = 'Country', encoding = 'ISO-8859-1').
   - Select ['Canada', 'United States', 'China', 'Australia'] and compare their population growth. Use the following formula:

   $$growth = \frac{1}{period} \times \frac{(value_e - value_s) * 100}{value_e}$$

where $value_e$ and $value_s$ are the population values at the start and end of the given period.

- Compute the sum and mean of CO2 emission for the same countries.
- Does an increase in population lead to an increase in the CO2 emission for all of these countries?
- Find the top 10 emitting (per capita) countries in each region for 2010.
- Is there any interesting trend in these countries with regard to their IncomeGroup?