Practical Exercise 7 -- Fun and games with OxGarage

James Cummings -- @jamescummings   *2014-11-13*

1   Learning Outcomes

If you successfully complete this exercise you should be able to:

- Transform a TEI XML file to a Word Docx or PDF

- Transform a Word Docx to TEI XML

- Use OxGarage for transforming files

- Open a Word Docx file in oXygen

2   Summary

In this exercise we take some files and use OxGarage to convert them from one format to another, and look at the results in oXygen.

3   Starting Up

In this case we're starting with two sample files (available from the http://tei.it.ox.ac.uk/Talks/2014-11-warsaw/ website that we have created for you. Download the files 'A65093.xml' and 'convertMe.docx'.

4   A65093.xml

This file contains a The works of an antiquary Polidore Virgil. It is from the Early English Books Online - Text Creation Partnership (EEBO-TCP). Glance through some of the file and look at the sort of markup contained in it. It is fairly minimal and structural markup. Note that some things are marked as highlighted (`<hi>`), and that it is divided into chapters, and that way down near the end there is a list that forms an index. We chose this file mainly to give you a fairly long text to convert.

5   Converting using OxGarage

- Go to OxGarage at `http://oxgarage.oucs.ox.ac.uk:8080/ege-webclient/` or if this is down/busy/not-working try `http://www.tei-c.org/oxgarage/`

- In the 'Documents' section of Oxgarage, select 'TEI P5 XML Document' as the input and 'PDF' as the output.

- Click 'Choose File' and find your downloaded copy of 'A65093.xml', then click the 'Convert' button.

- Look at the result. Did things transform as you expected them?

- Refresh the web page and try the same thing again but select 'Microsoft Word (.docx)' as the target format. Compare how this transformed to .docx with the PDF.

- Try some other outputs. Are some more successful than others?

- Use some (well-formed and valid) TEI P5 XML files you have worked on during the workshop as the input and compare their resulting transformations to other formats.

Clearly the underlying conversions could use some improvements if you were writing them for this file in specific. It is possible to create profiles and additional conversions that work on top of the basic conversions provided by the TEI.

6   convertMe.docx

This file contains a recipe for Chicken Biryani. It contains an image, a set of ingredients as separate paragraphs, and a list of steps.

- Open the 'convertMe.docx' file in MS Word (or OpenOffice / LibreOffice) or similar and have a look at the file.

- Having closed it, without saving any changes, upload it to OxGarage (as above) but choose 'Microsoft Word (.docx)' as the input and 'TEI P5 XML Document' as the output.

- OxGarage should produce a .zip file which contains an XML file and an image in a folder. Extract the contents of the .zip file somewhere convenient, and open the XML file in oXygen. How does the markup look to you? Could it be improved?

- Open the 'convertMe.docx' in your wordprocessor again. Try changing some of the content. (e.g. make the title of the file 'Heading 1' style, and the 'Ingredients' / 'Method' as 'Heading 2' style. Maybe make the list of Ingredients a bulleted list.) Repeat the conversion of this to TEI and see if better TEI XML is produced.

7   Word files directly in oXygen

You may open MS Word .docx files directly in oXygen if you want, because .docx files are really just .zip files renamed with a bunch of XML files inside.

- Open the file 'convertMe.docx' in oXygen. This will open an 'Archive Browser', where 'word' is one of the folders, and inside that you should find 'document.xml', double-click on it to open it.

- Look through the file (ignoring that even MS Word doesn't follow the published rules for .docx and so the file has a couple errors). We hope you'll agree that the TEI is at least easier to read!

- You could try changing some of the text (but don't change the markup), and after saving see the result in your wordprocessor by reloading the file there.

- Opening Word files directly in oXygen is something you probably don't want to do much...but it is one way of sometimes fixing corrupted and invalid Word files. (If the underlying XML markup is the problem.)