

Exercise 7: Customising the TEI

James Cummings

28 January 2014

1 Exercise 7: Customising the TEI

1.1 Learning Outcomes

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

- Analyse the TEI elements, attributes and values you need for your TEI XML document
- Tailor a TEI schema to your TEI XML file in Roma
- Use a different schema in oXygen
- Generate human-readable specifications of your TEI schema in Roma
- Set the value of existing attributes
- Be aware of the underlying TEI ODD XML format

1.2 Summary

In this exercise we will customise the TEI to remove those elements we do not think we'll use. In order to customise a TEI schema you need to know which elements you want to use, and which you don't, which sometimes involves a lengthy document analysis process. In our case we'll shortcut that by telling you what to include or exclude. You will learn to create a new schema, and download and use it in oXygen. You'll learn how to constrain the acceptable values for an attribute, and require its presence. You'll have a look at the underlying TEI ODD XML format which enables this customisation.

1.3 Starting Up

Load up the file from the previous exercise, or get it from 'spoilers/exercise-06.xml' in oXygen and save it under a new name where you've put your other files. Open up a web browser and go to <http://www.tei-c.org/Roma/>. (There is also a development version of this at <http://tei.it.ox.ac.uk/Roma/>.)

1.4 Your Current Schema

oXygen already knows about the TEI, it comes bundled with an open source TEI Framework (oxygen-tei) that helps it understand how TEI files are meant to work.

1. In oXygen with your file loaded, move the cursor to just inside a paragraph after the opening `<p>` just before where it says 'In most humble wise'.
2. If you type a '`<`' at this point, as you know, oXygen will give you a drop down list of all the elements allowed inside a `<p>`.
3. Scroll down the list of elements, referring to the pop-up tooltip if you want to know what the elements are for. Notice such elements as `<address>`, `<camera>`, `<incident>`, `<metamark>`, and `<notatedMusic>`.

4. Hit escape to leave the drop down menu and delete the '<' that you had added.
5. You certainly have a lot of choices for elements you can add here! But in any project it is unlikely that you are going to want all those choices. Also, increased choice of what elements to add can lead to greater human error and inconsistency, and we don't want that!

1.5 Roma: Starting a New Schema

Roma enables you to customise the TEI schema and remove those bits you are not going to use.

1. Go to <http://www.tei-c.org/Roma/> in your browser and note that you are given five options from which to start:
 - (a) Build up: this allows you to create a new customisation by adding elements and modules to the smallest recommended schema
 - (b) Reduce: this allows you to create a new customisation by removing elements and modules from the full `tei_all` (largest) schema
 - (c) Create a new customization from a template: this enables you to start from a convenient point for certain customizations
 - (d) Use or modify an existing TEI-defined customization: this allows you to start from one of the major customizations the TEI supports
 - (e) Upload a customization: this enables you to start from any customization file you had previously saved
2. In our case, let's start by choosing 'reduce', and clicking 'start'.
3. Set your parameters, change the following things:
 - (a) Title: 'TEI with maximal setup' is kind of boring, why not call it something like 'My special TEI customisation'.
 - (b) Filename: change 'tei_all' for something like 'myTEI' (don't include spaces).
 - (c) Author name: Change this to your name.
 - (d) You can leave the description as it is for now.
4. Click 'Save' at the bottom of the page. Notice how the box in the upper right tells you which customisation you are working on.

1.6 Adding and Deleting Modules

Modules are groupings of TEI elements for structural or semantic reasons. For example there is a 'dictionary' module which contains most of the elements needed for writing dictionaries. If you are not writing/digitising a dictionary, you probably don't need that module. Below is a list of all the TEI modules:

Table 1: List of TEI Modules

analysis	Simple analytic mechanisms
certainty	Certainty and uncertainty
core	Elements common to all TEI documents
corpus	Corpus texts

List of TEI Modules(cont.)

dictionaries	Dictionaries
drama	Performance texts
figures	Tables, formulæ, notated music, and figures
gaiji	Character and glyph documentation
header	The TEI Header
iso-fs	Feature structures
linking	Linking, segmentation and alignment
msdescription	Manuscript Description
namesdates	Names and dates
nets	Graphs, networks, and trees
spoken	Transcribed Speech
tagdocs	Documentation of TEI modules
textcrit	Critical Apparatus
textstructure	Default text structure
transcr	Transcription of primary sources
verse	Verse structures

1. Click on the 'Modules' tab to go to the page that allows you to add/delete modules from your schema.
2. Notice that because we've started with a 'maximal' schema, the list of selected modules on the right is completely the same as the list of TEI modules on the left.
3. Click 'remove' next to 'analysis' on the right-hand side. Note that it vanishes from this list, but remains on the left-hand side where you could add it back if you wanted it.
4. Remove 'analysis', 'certainty', 'corpus', 'dictionaries', 'drama', 'figures', 'gaiji', 'iso-fs', 'linking', 'nets', 'spoken', 'textcrit', 'verse', and 'tagdocs'!
5. Well! With removing that many maybe we should have started by building up instead of reducing down? You should be left with: 'tei' (you can't remove this one in Roma), 'core', 'header', 'msdescription', 'namesdates', 'textstructure', and 'transcr'. Why do you think we have left these modules?

1.7 Including or Excluding Elements

We have shrunk down the TEI to just a few modules, but those modules contain elements that we don't want.

1. Click on 'core' (note: not 'remove' but the word 'core') on the right-hand side. This should take you to a page listing all of the elements in the 'core' module.
2. Each row of this table has:
 - (a) the element
 - (b) whether it is Included or Excluded
 - (c) the name being used for the element
 - (d) a question mark linking to the reference page for this element

- (e) a description of the element
 - (f) a link to change its attributes
3. It is possible to Include or Exclude all the elements by clicking this word in the table header.
 4. From 'core' exclude the following elements: 'addrLine', 'address', 'analytic', 'biblStruct', 'binaryObject', 'distinct', 'divGen', 'gb', 'headItem', 'headLabel', 'imprint', 'index', 'listBibl', 'measure', 'measureGrp', 'meeting', 'mentioned', 'monogr', 'postBox', 'postCode', 'relatedItem', 'rs', 'said', 'series', 'sp', 'speaker', 'stage', 'street', 'teiCorpus', 'term', 'textLang', and 'time'.
 5. Wow! That's a lot less elements in your TEI schema. Remember to click 'Save' at the bottom of the page!
 6. We could go through to each of the other modules removing elements from there, but you get the idea. In a real life situation you would work through carefully only including elements that you really needed. The tighter your schema, the more consistent your data!

1.8 Saving Your Schema

1. If you click on the 'Schema' tab you will see a drop down menu listing various schema formats to generate. The TEI uses a meta-schema format of its own called ODD which allows it to generate these different formats.
2. Generate a schema either in Relax NG Compact Syntax, or Relax NG XML Syntax. These really are the best choices.
3. When you click 'generate' your browser should automatically download the schema file. Find wherever it has saved it, and move the schema file (not copy, move) to the place you have saved the 'exercise-06.xml' (or whatever you named it) file. They should be in the same directory.
4. Do not close down your browser window or you'll have to do that all again.

1.9 Associating Your Schema in oXygen

oXygen has been using the tei_all schema by default because it recognises (from the TEI element in the TEI namespace) what kind of files we have been creating.

1. Go to oXygen and the file you have previously loaded.
2. With this file open go to the 'Document' -> 'Schema' menu and note the icon next to 'Associate Schema'. This icon should also be on your oXygen toolbar. Click either the icon, or 'Associate Schema'.
3. Click on the little folder icon in order to 'Browse for local file'. Find the schema file you saved earlier, select it, and then click 'OK' when back in the oXygen dialog box.
4. When you click 'Ok' then oXygen, depending on your version, should add a line that looks something like this:

```
<?xml-model href="myTEI.rng" type="application/xml"
schematypens="http://relaxng.org/ns/structure/1.0"?>
```

at the top of your file.

1.10 Trying It Out

Remember those elements like 'address' and 'camera' that you could add within a paragraph?

1. Go to somewhere just after a `<p>` opening tag, and insert an '`<`' to get a drop down list from oXygen.
2. Are any of the elements you excluded available? No? Good! If they are, then chances are you didn't click 'Save' after Including/Excluding them, go back and do it again!

1.11 Constraining the @type Attribute on `<div>`

Removing elements is all well and good and is the first step in customising your schema, but we want to do more. Let's customise the @type attribute on `<div>` to only allow certain values.

1. Go back to Roma in your browser (hopefully you didn't shut it and lose all your work?)
2. Click on the 'Modules' tab.
3. Click on the 'textstructure' module name.
4. On the row containing 'div' click on 'Change Attributes' on the far right-hand side.
5. This should take you to a page listing all the possible attributes on `<div>`. This is also where you would include/exclude use of those attributes if we wanted to change that.
6. Scroll down to 'type' and click on it. This should take you to a page allowing you to set various options for the @type attribute. Set them as follows:
 - (a) Is it optional? This allows us to control whether the attribute is required or not. Let's make our @type attribute required, so click 'no' it is not optional.
 - (b) Contents This would allow us to change what type of datatype is allowed and how many times it should appear. Let's leave that just as it is as 'Text'.
 - (c) Default value would allow us to set a default value for the attribute if you didn't supply one. Let's force ourselves to supply one and so leave this blank.
 - (d) Closed list? enables us to say whether our list of values is fixed, or merely a suggestion. Let's be rigorous and say that it is a closed list. Answer yes!
 - (e) List of values is where we give the values we want to supply to the schema as valid values for the @type attribute on `<div>`. We give this as a comma-separated list. So write in: `prose,verse,drama,chapter,somethingElse`.
 - (f) Description allows us to change the description of this attribute. Add the phrase 'Our modified type attribute' to the start of the description.
7. Click 'Save' at the bottom of the page.

1.12 Trying It Out Again

Let's go try out the changes we made. You know how to do this now:

1. Click on the 'Schema' tab.
2. Choose one of the Relax NG formats from the drop down list.
3. Click 'Generate'.

4. Find where the file has downloaded it and move it over the previous version you had in the same folder as your file.
5. Do not close down your browser!
6. Go back to oXygen, and your file, and go to the 'Document' -> 'Validate' -> 'Reset Cache and Validate' menu item.
7. Your document should validate fine and you should have a happy green square.
8. Go to the first `<div>` tag in the document that looks like `<div type="prose">` and change it to be just `<div>`.
9. Your document should not be valid. You should have an angry red square. If it is still valid 'Reset Cache and Validate' again, and ensure that it is pointing to the correct schema. The error message it should be providing is that that element 'div' missing required attribute 'type' or something similar.
10. Put your cursor immediately after the 'v' in `<div>` and press space. oXygen should provide a drop down list of attributes available on `<div>`. Scroll down until you find `@type` and note that it is in bold. This is because we made it required.
11. Select `@type` and notice that oXygen gives you another drop down list of the possible values. This is because we provided the values and said that this was a closed value list.
12. Choose one of the values, perhaps 'prose'. Your document should again be valid and have a happy green square.

1.13 Saving Your Customisation

This is great, but what if you want to save your customisation, and come back later to do more work?

1. Go back to your web browser and click on the 'Save Customization' tab.
2. Your browser should automatically start downloading an XML file. Move it to somewhere convenient, for example where you put the schema.
3. Do not shut your web browser yet!
4. This is the file that you could upload when going to the 'New' tab on Roma (the very first page with the four choices), if you had selected 'Open existing customization'. (Don't do this now though!)
5. Open this XML file in oXygen. It might not be formatted or indented properly. If not go to the 'Document' -> 'Source' -> 'Format and Indent' menu, or click the Format and Indent icon on the toolbar, or press 'control-shift-p'.
6. Read through the file to get a sense of how it relates to your customisation. Note how `<moduleRef>` includes those modules you have asked for, and how the 'core' module is included except for the list of attributes you excluded.
7. Look at the `<elementSpec>` for `<div>` and see how we've changed it.
8. Note that this file is a TEI file just like the ones you've been editing, it just uses special elements from the 'tagdocs' module.

1.14 Generating Reference Documentation

Roma does not only generate schemas, but also customised reference documentation.

1. If you return to your web browser and click on the 'Documentation' tab.
2. Choose HTML web page from the drop down menu and click 'Generate'.
3. If your browser has downloaded the file, instead of opening it, open the saved file with your web browser.
4. You should get a web page starting with a table of contents listing the elements. Scroll down and click on <div>.
5. Notice that this has the @type attribute as required, and lists the legal values. Notice, however, that the example has not changed and has values that aren't in your list.
6. Try generating some PDF documentation as well. Which do you prefer?

1.15 More About Roma

Roma the web front-end is a dated web interface to both a command line script and the OxGarage web service. (<http://www.tei-c.org/oxgarage/>) When you generated the documentation this used OxGarage behind the scenes and you didn't even notice!

Some people write their TEI ODD customisation files entirely in XML and do not use the Roma web interface at all. There are a number of things that the Roma web interface can't do which the TEI ODD language underneath is capable of. Notice, for example, that you weren't able to provide descriptions of each of the attribute values you entered for @type? You can do that in the underlying XML. Some people do a combination of both Roma and hand editing.

There is also a 'Sanity Checker' tab... click that and find out what happens! (It might warn about the element <term> being used in <keywords> but not being defined. That is fine!)

1.16 Self-Assessment

Check if you understand some of the core principles of this exercise by answering the following questions:

- What is Roma?
- How do you add and remove TEI modules using Roma?
- How do you include/exclude individual elements using Roma?
- How can you change attributes using Roma?
- Is it possible to save your customisation in Roma?
- What kinds of documentation can you generate in Roma?
- What kinds of schemas can you generate in Roma?
- What does an underlying TEI ODD customisation file look like? Is it a TEI file like the ones you've been working with?

1.17 Further Reading

Consider reading more about TEI ODD at:

- <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/USE.html#IM>.
- and also the Documentation Elements chapter at <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TD.html>.
- See also, <http://tbe.kantl.be/TBE/modules/TBED08v00.htm>.