

Getting started with Emacs

Lou Burnard Date: (revised 23 Feb 03)

1 Using Emacs to edit XML

GNU Emacs, in its SGML/XML mode, is a very attractive option for those wishing to produce high quality XML-encoded documents on a limited (or nonexistent) budget. The software itself is genuinely ubiquitous, very widely understood, documented in great detail, and comes at a price that every academic project can afford (i.e. zero). It may not be as flashy as some XML editors, but it can still be configured to make entering valid XML text very simple. In this exercise, we'll use a version that has been customised to simplify data entry of TEI conformant documents of any kind. We won't try to explain all about emacs as there are dozens of books and web pages on the subject (as a taster, there is a very good tutorial specifically on the XML customization of emacs), and emacs also comes with its own built in help system and tutorial.

2 Making a new TEI document with emacs

Start Emacs by clicking on the Emacs icon. On the first screen that appears you will see the following mysterious suggestion:

```
If you want to create a file, visit that file with C-x C-f, then enter the text in that file's own buffer.
```

Let's first explain the notation: C-x C-f means

- 1. hold down the CTRL key
- 2. without releasing the CTRL key, hit the X key once
- 3. release both keys and
- 4. hold down the CTRL key (again)
- 5. without releasing the CTRL key, hit the F key once
- 6. release both keys

It's probably a good idea to practice this manoeuvre, if you're not used to using the keyboard. Emacs dates from pre-mouse days and some of its delights are accessible only via the keyboard. Most of the commands available on the menus at the top of the screen are also accessible using the keyboard, and often once you've learned the keyboard combination you'll find it quicker to use that than to fiddle around with the mouse.

Whether you are going to create a new file, or open an existing one, you proceed in the same way, either by selecting "Open File" from the Files menu, or by typing C-x C-f. On a Windows system, the first method will lead to the usual Windows open dialogue. Navigate your way to your work directory, and create a file called punch.xml.

The second will cause the cursor to disappear from the screen window and move to the bottom of the screen into a tiny window which is called in emacs-speak the *minibuffer*. This where you give emacs commands that require keyboard input. Emacs insists that you supply a name (even for a newly-created file), which is why the cursor has moved to the minibuffer. You can use the arrow or backspace keys to edit the text emacs is suggesting, and you can type in whatever you like -- but you must type something. For this part of the tutorial, please enter h:/punch.xml, and then press RETURN (if you get into a mess and want to start again, you can bail out of almost any emacs command by typing C-g).

Emacs now opens a new window, and you will see that the list of options available across the top of the screen has changed to indicate that you are now editing an XML file. If you've never used emacs before, you'll be relieved to learn that most of the keys behave in the way you'd expect (try moving around the

Date:

screen with the arrow keys, PgUp Home, backspace, etc.). The mouse can also be used in the usual way to select text, scroll up and down, etc. Selected text can be deleted with the Del key, and most basic editing functions are available from the Edit menu. You may like to experiment with some of the commands there before proceeding. Note that a few key combinations don't do what you expect: C-c and C-v don't cut or paste because they are used by emacs for other purposes.

In this exercise, we'll make a short TEI header for our file, and then build up a complete TEI document that can be properly validated, and which can make use of the XML-specific features of emacs.

- From the DTD menu (second from the right) select the second item, Insert DTD. A sub menu appears, offering you a choice of predefined DTDs. Choose the option (XML) TEI Lite.
- Because emacs already knows about this DTD, it can do a lot of the work for you, as you will now see. It will begin by inserting something like the following lines at the top of your file: <?xml version="1.0"?><?xml-stylesheet type="text/css" href="teixlite.css"?>

```
<!DOCTYPE TEI.2 PUBLIC "-//TEI//DTD TEI Lite XML ver. 1//EN"
```

```
"c:/dtds/xml/tei/teixlite.dtd">
```

These lines you will recognize, probably, as the incantantions necessary to state (a) the document following is an XML document (b) it may be formatted using the teixlite.css stylesheet, and (c) it conforms to the TEI Lite XML DTD. Move the cursor down to follow the last of these lines.

- You could just start typing in XML at this point. But an easier way is to go to the Markup Menu (fourth from the right) and choose Insert Element. As you are at the start of your new document, only one element is legal: <TEI.2>. Choose it, and see what emacs does to help you.
- Wherever it can, emacs puts in the tags that are required by the DTD automatically. Where the DTD requires something, but gives you a choice, emacs puts in a comment (in red) to show you what the choices are. According to the teixlite dtd you must supply for any document
 - a title (that's where the cursor is)
 - a publication statement
 - a source description
 - and, within the text itself, either a <body> or a <group> element
- Type in a title for your new document ("An extract from Punch" for example) and then move the cursor to the start of the first red comment.
- Choose Insert Element from the Markup menu again: you will see that the choice of elements available is now different. We suggest you choose here. Within the , you will find a very large number of elements is available: for simplicity we suggest though that you just type in a piece of text such as Unpublished exercise.
- Repeat the same procedure for the <sourceDesc> element.

Explore what options are available with other elements. Think about the metadata which it would be helpful to supply for the Punch sample pages.

Finally, insert one of the small XML documents you made last week at the right place in the document (inside the <body> element) by putting the cursor there and selecting Insert file from the Files menu, or typing C-x i. Either way, you'll be prompted to type the filename (try drama.xml) in the minibuffer, and press return, as usual.

Now that you have a complete XML document, you should save it and validate it

- From the XML menu (6th from left), choose Validate. Or type C-c C-v. This will run the nsgmls XML parser to check the file you are editing.
- The command line used appears in the minibuffer at the bottom of the screen: press RETURN to confirm.
- In order to show you the output from the parser, emacs will split the screen in two. You will see any error messages in the new window. Try to fix the errors, and repeat until you have a valid document.

Which messages you see depends, of course, on the validity of your file. Here are some typical examples, with a brief explanation:

- nsgmls:punch.xml:12:20:E: end tag for "publicationStmt" which is not finished The parser has reached character position 20 of line 12 in file punch.xml, where it found an endtag for the publication statement, but has not yet found some content which the DTD expects in a valid publication statement
- nsgmls:punch.xml:16:0:E: character data is not allowed here The parser has found some untagged text at the start of line 16, but it was expecting to find the start of an element at this point.

Often you will find that fixing one error causes many other apparent errors to disappear. Remember that the messages are meant to be helpful: you shouldn't take them personally!

You should be able put enough tagging into the drama.xml file to make it valid, but just in case you get stuck, we've prepared a valid version of the document for you: it's called drama-valid.xml..

3 Editing a valid document

Satisfying though it is to have a *valid* document, we are still some way from having a *truthful* one! For example, the TEI <sp> element is meant to mark a single speech, not (as here) several of them. If we don't separate the speeches out properly, we won't be able to process them properly. Fortunately, emacs in xml mode can help simplify this task. Here are some suggestions:

- Use the mouse to put the cursor at the end of a speech and select Insert end tag from the Markup menu: only one end-tag is possible. (You could achieve the same effect by selecting End current element, or typing c-c /). Repeat, to close the <sp> element as well the Then move the cursor to the start of the next line, and select Insert start-tag from the markup menu: choose the right start-tag, and proceed.
- Put the cursor at the end of a speech and type C-c RETURN. Each such step will split the current element, in the same way as you saw with Xmetal last week. Take care though -- you will wind up with the speaker prefixes inside the paragraphs (use the emacs commands on the Edit menu to cut and paste)
- Or you could just type the tags in...

As you've noticed, emacs doesn't require you always to insert complete elements (both start and end tags), so it's possible for your document to become invalid. But any emacs command that operates on elements, rather than just tags, won't work unless your document is valid.

If you get tired of splitting speeches, we suggest you tag a few other things, such as the stage directions or emphasis. As long as your document is valid, you can use the Tag Region command on the Markup menu to do this. Simply select the name or string you want to tag using the mouse, and then choose Tag Region. A menu of available elements appears.

When you need to insert an attribute value, you can type it in directly, inside the tag, or you can let emacs put it in for you. For example, it would be useful to supply the who attribute for each speech. Put the cursor inside the first <sp> start-tag. You could just edit the tag directly by typing in the characters "who="lady"" for example. Or, more safely, you could choose Insert Attribute from the Markup menu: select who from the submenu, and then enter the string lady in the 'minibuffer' at the bottom of the screen. Press return to add it to the document.

There is a further range of useful XML-specific facilities on the Modify menu: for example

Kill element delete the current element entirely

Untag element remove the tags for the current element

Change element name change the tags for the current element

- Fill element fill and indent text content of the current element, removing redundant white space (you can control the amount of indentation using the File Options command on the SGML menu).
- Edit attributes edit the attributes of the current element

Most of these have keyboard shortcuts, which are given on the menu.

You can use the Next trouble spot (C-c C-o) command on the Move menu to move through the document from error to error, fixing them as you go. Check to see when you have fixed all the errors by running C-c C-v again. Remember to save the document (C-x s) before validating it again.

If your document gets into a mess, you can always undo your last action, by choosing Undo from the Edit Menu, type C-_ ,or C-x u. If you get irretrievably confused, try C-g to abandon the current editing commands. And, yes, if all else fails, we have prepared a more fully tagged version of the file for you... it's called drama-true.xml