

Back to the wordprocessor

It's been quite a while since I have looked at word processing here. Over the next couple of weeks I'm going to revisit some of the ongoing themes I cover on this blog about document formats and [Scholarly HTML](#) and so on, working with my colleague Ron "If The Encodings Don't Get You The Namespaces Will" Ward, a veteran word processor wrangler, the man who wrote most of the rendering and conversion code in the ICE application. (I'm still working on the [metadata stores](#) thing, too).

So, in no particular order, some things we're going to look at.

1. I'll come back to the wonderful world of Custom XML in Word 2007, [which may be a ex-feature](#), and the work that Microsoft Research did on various authoring plugins that used it, and try to find out what's happening there. We have a visitor coming this week who worked with MS Research on their ontology plugin. More about that soon.
2. I'm going to revisit the Word 2007 interface, which frankly still frightens me, I dread that process of hunting through weird interface widgets for something I used to know how to do in muscle memory.¹
3. Interop between Microsoft Word 2007 and Openoffice across **.doc**, **.docx** and **.odt**. This is not just a technical issue, there's politics involved. For instance I think, [Sun's](#) evangelical policy used to be "Over our dead body will Writer save into that filthy Microsoft .docx format, although we will let you read it in, in order that your documents might be saved." (That's *saved* as in saved from eternal damnation *as well as* saved in the one true open format). Dear [Larry](#), we have the dead body part, can we have Save asdocx now²?

Closer to home there are political issues at USQ with ICE and how it works with some classes of documents. Over in the Faculty of Engineering I gather people are being forced to use ICE for maths-heavy courses, Bron Chandler has been working hard to help the course-maintainers in engineering, to find an acceptable way for them to use Word or Writer plus ICE but there *are* limits to how far we can push these word processors for heavy duty technical documents. I'll say it again: I don't think it makes sense to mandate the use of a tool like ICE at a university; we should talk about performance based standards for materials and remain pragmatic; if we can't find a cheap way to make HTML versions of maths-heavy materials then we may have to settle for PDF and we may have to let people use LaTeX, or DocBook or whatever (I've never heard of one wanting to use DocBook but there are plenty of LaTeX enthusiasts).

4. Ron is going to look at how we might be able to use Word instead of or in addition to OpenOffice.org in ICE. We picked OpenOffice.org as the main engine-room for ICE several years ago because it is available cross-platform but what if we could do one or more of the following on Windows?
 - Use Word to generate PDF – should improve the WYSIWYG fidelity (Ron's already got this automated, although not hooked in to ICE yet.)
 - Use Word rather than OpenOffice to render the image parts of a page, again to improve the way things look. ICE uses a two-part process to convert to HTML. It converts OpenDocument format XML into HTML, but to get images such as charts or equations, it also calls OpenOffice's Save as HTML feature (initial experiments look promising).
 - Maybe replace OpenOffice.org altogether on Windows by automating it to save as .odt behind the scenes. (If that works, there is still the issue of how to make books out of individual course documents. At them moment this I a very complex process which automates Writer, as the master

1 Look, I'm not afraid of learning new things, why at the moment I'm more or less learning to play the mandolin which is tuned in 5ths rather than 4ths like most of a guitar – that's a whole lot of new fingerings, but it makes *sense* and it's fun.

2 Yes I can save as .docx here in Ubuntu, but that's a special build with some Novell code in it, I believe.

document feature doesn't work for what we want.)

- And if the above fails, maybe rewrite the ICE renderer for Office Open XML rather than Open Document Format. Big job we'd rather not have to resource.

Along the way, this is going to force me to spend some time in Windows for the first time in a long time, guess I'll have a chance to learn about Windows 7.

Copyright Peter Sefton, 2010. Licensed under Creative Commons Attribution-Share Alike 2.5 Australia.
<<http://creativecommons.org/licenses/by-sa/2.5/au/>>



This post was written in OpenOffice.org, using templates and tools provided by the [Integrated Content Environment](#) project and published to WordPress using [The Fascinator](#).