More ideas about online and offline word processor integration - is anybody listening?

Via Glyn Moody who doesn't want to say he told us so I see that Adobe is discontinuing support for Flashpaper, a proprietary Adobe (via Macromedia) technology for disseminating documents online. This means that anyone who has put stuff in there now has to migrate all their stuff to some other format. That's what you get for using technology that's controlled by a single vendor.

That reminded me that I had this piece I've been working on about Adobe Buzzword, another Adobe proprietary document format.

Following my last post on Buzzword, I had an email from Tad Staley at Adobe which seemed encouraging:

> You had an interesting point about exporting named styles to Word. By this, I assume you mean that we create a handful of styles that correspond to Buzzword fonts and paragraph settings, and use them within the .doc or .docx file we create on exporting? This would then allow us to "round trip" the document better from Buzzword to Word and back again.
>
> We'd like to hear any other thoughts you have with respect to styles - as I said, we're working on them now, so the timing is good.

So I drafted something along the lines of what you're about to read and sent it off to Tad. It was pretty clear from Tad's reply that Adobe are not thinking along the same lines as me at all . They don't see it as important to be able to interchange with other word processors because they're going to make theirs broadly available and they don't care much for HTML because they care too much about controlling the fine details of document presentation. What this means it they'd like you to use Buzzword / Flash / PDF to disseminate your work rather than an open web format – OK some PDF is a bit open but it is very much page oriented and much harder to integrate with other services than HTML. I think that's terribly short sighted and reduces considerably what people can do with their documents. Mashups and so on that are built for the open web would all have to be redone for the Buzzword world for example my geocoding example.

I'd be cautious of Buzz word out if I were you, because this format could easily go the way of Flashpaper.

Anyway here's the gist of what I sent to Tad at Adobe.

I think it would be a good idea for you to map Buzzword docs to a set of styles when you export to .doc or .docx- I'd like to see .odt as well. I'll go into specifics below but first some general comments.

There are so many issues with styles in word processors regarding styles, interop and HTML export it's a bit hard to summarize in a short email or blog post, but here are some of the main problems. It would be great for someone to get it right for once:

1. **No standard set of styles:** Nobody ships a rational set of styles by default - I'd be looking for something that covers headings (both numbered and not in the same document) lists, block-quotes, pre-format text; the list is actually very similar to the set of elements in HTML, which is no coincidence as that's a generic schema.

2. **Awful HTML export:** Word processors almost always try to reproduce whatever the user inputs in the way of formatting resulting in all sorts of crap in the HTML they output. Building a new product is a great opportunity to do it differently.

   (Buzzword's HTML isn't bad by comparison with some, within its current limitations. But really, you should fix the list nesting. I think the HTML model is silly too, but it is what it is.)

3. **No 'structure-only' mode:** Why not have UI mode where you can't do gratuitous formatting, only structure your document using headings, do lists and blockquotes etc and then choose from a menu of stylesheets? That is, turn off the font panel. This may have been hard to sell in the old days but now I think people would get it   particularly when they are writing for multiple media   where the same document could be published as both HTML and PDF. If you restrict users to a known style set then you can reliably change the presentation of their documents automatically. If not then you have problems. A couple of examples:

   i.   If people have chosen colours then you can't change the background colour of a page in case you have readability problems.

   ii. If you allow absolute indents (say 4cm) then you might not be able to reformat into multiple columns and still have the document look OK.

4. **Extreme confusion in the area of lists:** Both Word (and by extension OOXML), and Writer (and by extension ODF) have mind-blowingly crazy list models.

   - Word has paragraph styles to which you can attach list formatting, and it has named outlines (with one of the worst GUIs **ever** even before Word 2007 took it to new heights) AND it has list styles which came along circa Word 2003.

   - Writer has paragraph styles and list styles both of which can be applied independently, and lists are represented as a hierarchy in the file format, although he GUI gives almost no clues as to what the hierarchy actually is.

   In the ICE project we deal with this by automatically creating paragraph styles and list styles / named outlines and providing toolbars to apply both at once, resulting in much more stable, interoperable documents than you get if you leave users to deal with all this by themselves.

So here's what I would do if I had a chance to influence Buzzword, in addition to building in a standard kind of word processor style system.

Based on my observation of the behavior of the list formating Buzzword obviously has some notion of structure built into it even if it doesn't (yet) have headings. So lets look at what you could do with lists.

I think the Buzzword UI for lists is pretty cool – one thing I like is that lists stay connected. In most online editors if you change an item in the middle of a list into a plain paragraph and then back into a list item you get two disconnected lists, something that makes no structural or practical sense. Buzzword gets this right and makes sure that list items adjacent to each other are part of the same list.

Here's a test-list in Buzzword:
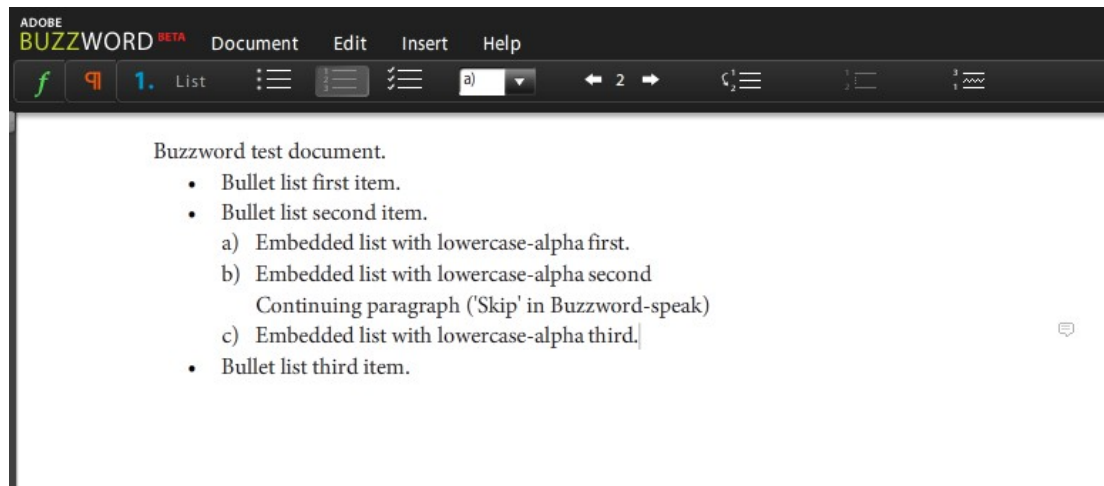
The UI is really slick – it actually understands the structure of the list so when you hit the promote (<-) and demote (->) buttons it does The Right Thing. My only quibble is the way it insists on all the items at a particular level being the same kind of list item even if they are not siblings.

Oh, and don't call the list level 'outline level' because in other word processors that term is used for the heading structure.

My proposal is that on export Buzzword should not just use formating it should create styles. As I mentioned before this is more complex than it needs to be, due to the legacy of gratuitous features in the target applications but it is doable.

Lets take the example of .odt export for use in the OpenOffice.org family of word processors. I'll use the ICE version of the style names, chosen for their brevity but you could use longer versions.

Here's the same test list embedded in this document which I'm writing using NeoOffice. The paragraph style names are shown in curly-braces at the end of each paragraph, behind the scenes my toolbar also applies a list style of the same name.

Buzzword test document. {p}

- Bullet list first item. {li1b}

- Bullet list second item. {li1b}

  a.   Embedded list with lowercase-alpha first. {li2a}

  b. Embedded list with lowercase-alpha second {li2a}

     Continuing paragraph ('Skip' in Buzzword-speak) {li2p}

  c. Embedded list with lowercase-alpha third.{li2a}

- Bullet list third item. {li1b}

When exporting, you could embed some macros that provide a buzzword-like interface via a toolbar. In ICE we have a toolbar which tries to Do The Right Thing (doesn't always succeed I have to admit, but we're getting there). We take a different approach from Buzzword's modal interface and re-use the same buttons in different contexts. So the promote button in a list will move your list item to the left and it should pick up the right list style by looking back through the document to see what is appropriate - whereas for a heading it would change the heading level in the document outline.

Why do we do this? It's all about interoperability. The styles mean that we can produce good HTML, and also move documents between Word and Writer pretty easily, correcting for the differences between their wacky, annoying, productivity-sapping list models. And we give users on both word processor the same toolbar running the same code.

One advantage for Adobe and their buzzword product would be the same – good interoperability with offline word processors. But there's another potential benefit, the same one I suggested to Google. Adobe could start 'infecting' documents with a benign structure virus. Lets see how this could work:

1. I draft a blog post like this one in Buzzword and send it via Buzzword's sharing feature to a colleague to add their contribution. I was going to say 'a paper' but Buzzword is a long way from ready for that.

2. My colleague doesn't want to sign up to yet another online service, and besides is going to be editing the document later at home, so chooses the option to download it as a Word document and saves it on a USB drive.

3. Later at home Word prompts to say that the document contains macros and should they be allowed to run? If no, then it's not the end of the world as we still have a Word document that can be re-imported to Buzzword later. If yes   then read on.

4. On opening the document, it's got a Buzzword-style or ICE-style toolbar, so my colleague is able to make some changes to the document without realizing that they are dealing with the styles that were added to the document automatically on export.

5. When the editing is done, they can save the document locally, but since there's a toolbar installed they can click the 'Return to sender' button and it gets automatically uploaded back into my Buzzword account via an inbox.

   Because they used the toolbar all the headings are set properly and the lists are nice and orderly.

   (If you don't understand why I'm going on about this go over to Google docs and try importing and exporting documents using OpenOffice.org Writer).

6. Later, if my colleague decides that they *did* like the Buzzword experience they can click the 'Install the Buzzword template' button and have the toolbar show up all the time. If they go further and sign up for an account then they can draft things in Buzzword and have them save automatically into Buzzword.

You can see how this could spread the Buzzword way of life not by replacing offline word processors but by providing a bridge into the online service. If the online way is better then people will naturally stop using their offline programs.

A couple of other things that would help drive the service:

- AtomPub support so you can post to your blog, both from the online service and from your word processor. ICE does this already.

- Simple web page publishing. At the moment Buzzword does HTML export in a Zip file why can't it just put the page up online for you?

- An import feature where when a user uploads an unstyled word processing document Buzzword gives it back with added styleage. (See the ongoing conversation I'm having with MJ Suhonos).

There would be a couple of ways for the online word processor vendors to approach this. One would be to work with the ICE team. As far as I know there is nobody else out there with our commitment to generic word processing based web and print content management.

The first mover would have an advantage and if it worked others would follow. The users would win.

Another would be to invent a proprietary set of styles and toolbars and go for more of a lockin effect. Might work. Wouldn't be so great for the users.

I am reminded writing this that all the recent activity on word processing standards hasn't changed things much for users. For complex documents, like business documents with embedded fields and so on interoperability between packages both online and offline is still really poor, and interoperability between word processing packages and the web is terrible. It's not about whether you're doing OOXML or ODF. It's about **what** you're doing with them.