HTML Email:
The Perils of Copy 'n Paste (Part I)

March 17, 2011   By Miller Ramos

While the title may sound like some sort of Hardy Boys mystery (albeit a genre-confused tale involving electronic mail and word processing), it's really only the preface for a more mundane happening involving two other lesser-known but widely-used siblings: Control-C and Control-V.

These days, copying and pasting text from one document to another is a pretty standard practice. Come on, admit it – you do it. I do it. Anyone who knows how to type does it. There's no shame here. It's one of the greatest editing shortcuts that technology has afforded us since the advent of sliced bread*. And given that content is written in and comes at us in a variety of formats, it's no wonder that we copy 'n paste with abandon. But take heed, gentle reader, for I'm about to tell you that there lies some modicum of danger in this very act.

..Er, maybe not life or death danger but danger involving that dastardly of miscreants – junk code. (We hatesssss it.)

Now, copy and paste does exactly as advertised: it makes a copy of source content from one location and inserts that copy at another location. Pretty straight-forward. Well, what you may not know is that in light of our ever expanding technological capabilities, copy and paste does a bit more than merely copy the content desired. It often copies the metadata that describes it, as well. In most cases this is a convenience as bold text stays bold from one doc to the next and bulleted lists aren't suddenly sentence fragments chained together when dropped into another article. Often, though, the metadata hitchhikers that come along for the copy and paste ride are the source of the junk code that appears in the very HTML that describes our email marketing newsletters.

Net Atlantic Email Marketing Blog

Best Practices dictate that junk code be eliminated from newsletters not only to keep file sizes down (less superfluous code = less information included), but to also prevent unexpected interactions with different email clients. Less code to wade through also makes for an easier time to navigate through it all and update it, even making the textual content portion more accessible as less gets in the way. In general, getting rid of junk code is just a good idea.

That said, have you seen any of these bad boys appear in your HTML email code?

<o:p></o:p>
If you see that code snippet (the 'p' may not always be there but the 'o:' part is what's worrisome) or any part of it, that means that the content made its way to the email by way of cutting 'n pasting from Microsoft Office. That there is specific Office code for their own proprietary (and application-side defined) styles. These should not be in the HTML if only because it isn't proper HTML but also because Microsoft applications (from IE to Outlook) may interpret that code in however fashion they may but other clients/applications may either ignore it, print it as text within the body of the email or, worse yet, do something completely unexpected as its attempt to interpret it. Often, they appear self-contained as in the example above and, because of that, can be easily removed altogether. You can get more on this from ckeditor's Developer Community.

class="Apple-style-span"
This snippet will appear inside any old tag but will mostly turn up in SPAN tags. This is indicative of the content being copied from an Apple product (namely, Safari) and pasted into the HTML editor, thereby inheriting the style class. This is a known bug with both Safari and Chrome as they both use the same HTML rendering engine (WebKit). Again, this should not be in

the code if only because it's superfluous but also for similar reasons stated in the topic above. While the style class isn't defined in the document itself, it can still be interpreted by those browsers (or email clients) that recognize the class because it's built into their respective applications already. If you see a SPAN tag and this is the only attribute defined in it, remove the entire SPAN tag altogether (including its associated /SPAN). If this appears within any other tag, just remove the class="Apple-style-span" part.

style="mso-bidi-font-weight:normal"
This is but one of many different versions of Microsoft Office generated CSS. The thing to look out for is the beginning part: "mso-". This should be gotten rid of for the same reasons as stated above. Remove this as you would the Apple-style-span ones.

*(Net Atlantic does not actually recommend sliced bread as an appropriate shortcut for text editing)