

How to Choose a Document File Format

W.M. Hinsch

8/8/2008

A common mistake made by computer users is to passively accept the file format offered to them by their word processing program. This is one of those areas where a little forethought can prevent a lot of grief later. One of the principles I try to share with others is the that of *using the least common denominator* when it comes to file formats. In other words, choose the least complex and most compatible document format wherever possible.

One factor that makes this more difficult is that we are easily cowed by the warning we receive from most programs when saving a document in anything but the default format. “Danger, Will Robinson! You just might lose a tiny bit of the incredibly important formatting you put in your document!” The key is to think it through. What formatting do I really need to keep, no matter what? What formatting is more flexible? And the toughest question, **“Is the value of my document found more in its style or its content?”**

Where formatting isn't as important and you can get away with plain text ([ASCII](#) or [Unicode/ .txt](#)), use it. This format will work fine a thousand years from now on any computer platform from [exaflop](#) supercomputers to cell phones, which in a thousand years will probably *be* exaflop computers. Think about it. Since the invention of printing, and later the typewriter, “plain text” has met the needs of everyone from authors to businessmen to clergy. These days, there is a glut of beautifully formatted garbage churned out using computers. If your document does not clearly communicate its message without extensive formatting, it should probably be rewritten.

Of course, text can have issues as well, but most of these are easily correctable. For instance, my e-mail is constrained to only use text format [for security purposes](#). The biggest issue I run into is the automatic line breaks inserted every 72 characters because of how email is transmitted, which brings up another point...

You probably already know this, but avoid using “hard” line breaks in text files. Type one continuous paragraph and only insert a return at the end of each paragraph. Text files saved with only paragraph breaks actually can work nicely in many cases since most programs (like Notepad) can handle word wrap automatically. Another advantage of plain text is the side effect of causing the user to think more about what they want to say than how they want to “decorate” their document.

One final note about text format. It can be a good idea to save important documents in text format in addition to a more advanced format. A document that took hours to write can be imported as text and reformatted in minutes. Again, it is the content which should be treated as valuable, not the formatting.

Before moving on to formatted documents, I would like to encourage you to learn the fundamental elements of typography. It is this understanding that will allow you to use *any* word processor or publishing program far more effectively. *The Elements of Typographic Style* by Robert Bringhurst is an excellent place to start. If you are an older computer user, this will help you understand the fundamental differences between “typing” and “publishing”. If you are a younger user, this will help you understand the roots of typography, including fonts, kerning, leading, serif vs. sans serif, proportion, special characters, etc.

Not wanting to be a hypocrite (for surely you have noticed that I have used formatting beyond plain text in this document), judicious formatting *can* make a document more readable and provide visual cues for the reader. [Rich Text Format](#), while owned as a standard by Microsoft is well documented, widely supported and of course, well supported by Microsoft programs.

Among open standards, [HTML](#) works quite well in most ways, but differences of versions and loose specifications can result in different interpretation of the formatting on different platforms. Either RTF or HTML may be written and read by almost any word processing program, and should be around for many, many years. They can handle all general formatting (fonts, styles, bullets, page setup, tables, etc.).

For more complex formatting, [Open Document Format \(ODF\)](#) will handle almost any formatting task with excellent cross-platform compatibility. This is better for documents with footnotes/endnotes, complex styles, TOC/index, outline formatting, etc.

[Portable Document Format \(PDF\)](#) is an open standard released by Adobe, and it has some advantages by allowing fonts to be embedded into the document and doing a more consistent job of recreating the exact appearance of the document across different platforms. It was created for this very purpose, and does a good job. It also has some security features. I often use PDF for documents I want to send, but not allow editing or copying. (There are ways around these limitations, but the effort makes it clear that it wasn't accidental or through a misunderstanding of the author's intent.)

Always, always, always *last on the list* are proprietary formats (native Word, Word Perfect, Lotus, etc.). These are designed to lock people into using one vendor because of the difficulty of translating years of documents into a different format. Particularly egregious are proprietary formats that don't work well – or at all – within the same company. I'm not sure if Microsoft ever got it worked out, but for many years their "Office" line of products was not file-compatible with their "Works" line of products. Ridiculous!

Proprietary formats generally do not add value to the final product. They tend to employ poor (and undocumented) underlying standards with huge compatibility and security issues. For instance, many companies routinely "mine" Word document resumes for **all** the editing changes made from the time the document was first created. Macros are now routinely disabled in office programs because of the capability of transmitting malicious code ([macro viruses](#)). In general, proprietary formats should be avoided except in cases where there is just no other choice. Even then documents which will be shared should be saved at least one full version behind the current version.

The good news is that most advanced programs allow the user to permanently set a preferred format for saving documents. For word processing documents, my preferred format is set to Rich Text Format. I can always change that when a document requires a different format. However, it helps prevents me from accidentally saving documents in a format I'll regret later.

In conclusion, spend a few seconds before clicking “save” the first time for each document. In addition to document name, intended audience, style, etc., make the decision on how you want to save your information a decision that is intentional, not accidental.

URL Cross References

ASCII

http://en.wikipedia.org/wiki/Plain_text

Unicode

<http://en.wikipedia.org/wiki/Unicode>

FLOPS

<http://en.wikipedia.org/wiki/FLOPS>

Secure E-Mail

<http://www.itsecurity.com/features/five-steps-email-security-092106/>

Rich Text Format

http://en.wikipedia.org/wiki/Rich_text_format

HTML

<http://en.wikipedia.org/wiki/HTML>

Open Document Format

http://en.wikipedia.org/wiki/Open_Document_Format

Portable Document Format

<http://www.adobe.com/products/acrobat/adobepdf.html>

Macro Viruses

[http://en.wikipedia.org/wiki/Macro_virus_\(computing\)](http://en.wikipedia.org/wiki/Macro_virus_(computing))

This document may be freely copied and distributed in its entirety. Please contact the [author](#) before modifying the document.