

ePMA Special Report:

eBook File Sizes Compared for Eleven Different Formats

To make this comparison, we took the Project Gutenberg plain text file of the book *The Metal Monster* by Abraham Merritt, stripped it of the PG legal disclaimer and such. Next we removed the linefeed/carriage returns from the end of each line, so paragraphs would flow for different formatting, and then added basic markup with HTML, to provide paragraph breaks and chapter headings. Except where noted, this HTML file was then used to create the eBook in other formats. No cover or other images were included. We hope to try a similar experiment with an illustrated volume in the near future, and will present those results in another report.

Obviously, this HTML file was very basic; if there were lots of bold or italics, tables or lists, or other formatting to be done, the HTML file could have been much larger. With just this basic formatting it was only 10K larger than the plain text file. As such, it produced a very plain eBook, but that was our goal for this project -- to keep things simple.

The results for nine of the formats fell neatly into three categories: Rocketbook, Mobi and Microsoft Ebook formats all produced small, compact files. The plain text, HTML and Embiid files were medium sized. MSWord, hiebook and Adobe Acrobat all produced large files. Of the remaining two formats, Microsoft's Wordpad produced a very large file, and the only Windows Executable program we tried (eBook Pack Express) produced a huge file. In defence of the executable file however, we must recognize that it was the only format that includes the reader within the same file.

The Rocketbook (.rb) format came in with the smallest file, 248K. The file was produced by the eRocket Librarian program (version 1.3.222) by NuvoMedia. This file was 52% the size of the plain text file.

Mobipocket (.prc) was next at 285K. This file was created using Mobipocket Publisher 3.0 Personal Edition from Mobipocket.Com. The Mobipocket file was 60% the size of the plain text file.

The Microsoft Ebook (.lit) file was also quite small, at 292K. The file was produced using Reader Works Standard 2.0 from OverDrive. This file was 61% the size of the plain text file.

The plain text file (.txt), at 477K was our starting point. If you were to compress this file using PKZip, it produces a file just 191K in size (40% of original), but of course the ZIP file can not be read directly, but has to expanded back to the original size for viewing -- it is convenient for transmitting or transporting the file however.

The HTML file (.htm) we created was 487K in size, but as mentioned above had only the most basic markup. It would be possible to have a considerably larger HTML file, especially if you used one of the more inefficient conversion programs to create it. We just used a simple program of our own creation, plus some hand coding, to come up with this basic HTML. This file was just slightly larger than the plain text, at 102% of the plain text file size.

The Embiid file (.ubk) at 489K was basically in the same range as HTML and plain text. It was created using the Embiid Writer 1.0 from Embiid. This was just 103% the size of the plain text file.

MSWord 2002 produced a large file (.doc) 620K in size. This was 130% the size of the plain text file.

The hiebook (.kml) was also large at 654K. It was created with the HiBuilder program (version 1.1.2e) from Korea ebook Inc. That is 137% the size of the plain text file.

The Adobe Acrobat file was a hefty 691K, or 145% the size of the plain text file. This eBook was formatted for reading on a computer screen using Acrobat Reader, but we also produced another formatted for the Acrobat Ebook Reader. There was no difference in file size. We used Adobe Acrobat version 4.02 -- according to Adobe's promotional material Acrobat 5.0 has better file compression, but we don't have that available for testing, so we used their free on-line conversion program. That, however, is limited to 100K, so we just converted on chapter, and compared it to the same chapter using Acrobat 4.02. The Acrobat 4.02 file was 36K, and the Acrobat 5.5 file 29K. So with the plain text version of that single chapter at 25K (100%), the html was 26K (104%), Acrobat 4.02 file at 36K (144%) and Acrobat 5.5 file at 29K (116%).

Microsoft Wordpad (ME Edition, which produces an MSWord 6.0 compatible file) produced a file (.doc) 952K in size from the original text. This text editor is used to edit HTML files, so it doesn't interpret the formatting, hence we used the text file instead. It increased the file size to double (200%) its original size.

We went to our favorite source for shareware programs and looked for eBook authoring software that creates Windows Executable files. The first two we tried were linked to websites no longer in existence. Finally we tried 'eBook Pack Express' which offered an impressive list of features. It produced a very large file, 2860K in size, six times (600%) the size of the original text file. Of course that includes the reader program, which in other formats is separate.

Format	Filesize	Kbytes	% of Plain Text
Rocket (rb)		248K	52%
Mobipocket (prc)		285K	60%
MS Ebook (lit)		292K	61%
Plain Text (txt)		477K	100%
Web Document (html)		487K	102%
Embiid (ubk)		489K	103%
MS Word 2002 (doc)		620K	130%
hiebook (kml)		654K	137%
Adobe Acrobat (pdf)		691K	145%
MS Wordpad (doc)		952K	200%
eBook Pack Express (exe)		2860K	600%