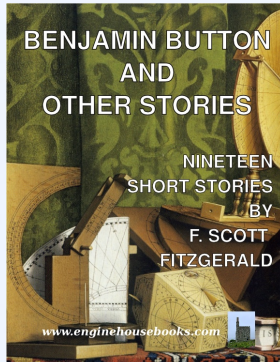




ENGINE HOUSE BOOKS

BOOK BITS #2



The release of our third e-book (*Benjamin Button and Other Stories*) seems like a good time to talk about some of the issues surrounding e-books. This will be the first of several *Book Bits* devoted to these issues.

Q: What exactly is an e-book?

A: For our purposes, an e-book is a computer file that contains the text of a book and which is intended to be viewed on some electronic device (rather than first being printed and possibly bound). The device might be a computer, an e-book reader, a phone or some other device capable of displaying text.

Q: What kind of file format is used for an e-book?

A: Many different formats are used, since many formats are capable of displaying text. The most common formats are listed in the accompanying table.

Q: Why are there so many different formats?

A: Several reasons. It is important to realize that e-books are really still an

E-BOOKS

MIME type	Description
text/plain	ASCII or Unicode text
text/html	Hypertext Markup Language
application/pdf	Portable Document Format
application/epub+zip	EPUB
application/x-mobipocket-ebook	Amazon Mobipocket
application/vnd.amazon.ebook	Amazon Kindle
application/x-bbeb-book	Broadband eBook (BBEB) (.lrf)

experimental medium (regardless of all the hype surrounding devices like Amazon's Kindle). As such, no one standard format for texts has yet emerged, and all of the formats in current use have advantages and disadvantages. No one format yet has all the capabilities that are likely to be necessary for the “ultimate” e-book format.

Q: What sort of capabilities are you talking about?

A: Principally, the ability to control things associated with the rendering of a page (for

example: type size; kerning; justification; font changes; images) and also what is sometimes termed the “control channel” (for example: is this device permitted to view this content?). We will defer a discussion of the issues related to the control channel to a later *Book Bits*.

There is a fundamental tension between the ability of the content author to determine the look of the text and the ability of the display device to control the rendering, and different file formats are designed with different philosophies as to which entity should ultimately be in control. As a general rule, formats that allow the creator to exercise a high degree of control are based on glyphs, whereas those intended to allow more flexibility are generally based on characters.

Q: What is the difference between a glyph and a character?

A: Taking them in reverse order (and simplifying slightly), a character is simply an instruction to render a particular letter on the display device. So, for example, a character-based format that contains the text “the ” will in essence send instructions to the device to display in sequence the four characters “t”, “h”, “e”, “space”. The file, however, contains no concept of what these individual characters actually look like. So they might be rendered, for example, in the font Times New Roman at size 10, or in New Courier at size 11 — the creator neither knows nor cares.

A glyph is an instruction to render a particular shape on the device (sometimes an explicit instruction as to the location of the shape is included). In this case the sequence “the ” causes the device to display individual shapes that happen to look like the individual characters “t”, “h” and “e” in some typeface; the actual shapes are contained within the file, not within the device. You might be wondering about the space; typically in glyph-based formats the location of the start of each character (or perhaps each word) is contained

in the file; the space is rendered simply by telling the device to start the next word in the correct location so that the correct space is left between the end of one word and the start of the next.

Q: Why does any of this matter?

A: Basically, if the creator of the content is in control, then he can be sure that the content renders in what he deems an acceptable manner. For example, he can be sure that: words are hyphenated correctly; inappropriate line breaks are avoided; paragraphs are indented correctly; all pages have the same length; widows and orphans are avoided; a particular typeface is used; and so on. The greater the extent to which the device is control of rendering, the greater the chance that the rendered text will likely suffer from the lack of some, or even all, of these.

Q: So why would a content creator not always use a glyph-based format?

A: Because it essentially confines the content to be displayed on a display of a particular size, and does not allow the user of the device to change the size of the displayed characters in a way that will preserve the details of the formatting. To put it simply, the creator essentially has to choose between: (a) making a file that will look good on a particular device but may be very inconvenient or even unsuited on other devices; and (b) making a file that can be displayed on a wide variety of devices but will likely look rather ugly on all of them. One can imagine ways in which this trade-off could be avoided (for example, by embedding in a single file several versions of the same text but using different glyphs) but to date none of these methods is in use.

Q: Which format is the most popular?

A: It's impossible to know. More texts are available in plain text than any other, but generally that format is not widely used for

dedicated e-book readers. Most books for the Amazon Kindle use the proprietary *application/vnd.amazon.ebook* format (usually called the AZR format). Many books for the Sony readers are in the *application/x-bbe-book* (often called the BBeB or LRF format; there is also an encrypted variant called LRX).

Q: What dedicated e-book readers are available?

A: Available readers vary from country to country. In the US, the one that most people know about is the Amazon Kindle (of which there are two models currently available), but there are several others. The Sony readers (two models) are popular. In total there are around a dozen readers currently on the market, and there is every reason to expect that number to increase. Naturally, each has its advantages and disadvantages. Each reader tends to support a different subset of the available formats.

Q: That makes it difficult to choose a reader, doesn't it?

A: Yes. All the variation and confusion also makes it difficult for e-book publishers. Content-channel issues merely complicate the situation even further.

Q: In what format does Engine House Books publish its e-books?

A: We publish e-books in the *application/pdf* format (often just called PDF). This is the same format as is widely used for documents published in electronic format intended for viewing on computers. This is a glyph-based format. We design our e-books for viewing on Sony readers; they can be viewed on other readers, and at sizes other than the intended size, but in those cases the book will look decidedly more ugly than was intended. We typeset our e-books explicitly for viewing on the Sony readers.

Q: Are you intending to support any other formats?

A: We are experimenting with the EPUB format. EPUB is a free and open standard for re-flowable digital text ("re-flowable" means that the device is in control of the location of the text on the page). However there currently exist no tools for automatically creating EPUB files from the files that we use for typesetting on paper; so it is unclear whether and when we will be able to release books in EPUB format.

Q: How many books do you intend to publish as e-books?

A: We are currently planning to release books in the Engine House Classics series as e-books, usually a month or so after release in paper.

Q: What about books not in the Engine House Classics series?

A: We have not yet made a decision about other books. Basically, we are dipping our toe in the water with the Classics series, and will see how that works out over the course of the coming months.

Q: How do you price your e-books?

A: It is our intention to charge a flat rate of \$2.50 for each e-book. A few texts (such as our Greek New Testament) will be offered at no charge.

Book Bits are occasional documents from Engine House Books featuring author interviews, industry insights and other information that might be of interest to our readers. They are published on an ad-hoc basis.