Monday, March 21, 2022 - Chapter-size

I realize that I am trying to accomplish several things in this project:-

(1) I thought to dig a project out of the mire; only two of (then) 70 sections were recorded, so I "rescued" the project from oblivion, as I thought.

(2) I wanted to make an impact, a statement, to members of LibriVox by showing what I could do with my time.

[1] (3) I wanted a solo project much more demanding than [Lawson](link) [2], so that I might learn more about the LibriVox machinery.

(4) I wanted to prepare myself for the load-balancing project of recording ten solo projects at once (of Australian-accented works)

(5) I wanted to improve my quantity and quality of Audacity 3.1.3 macros.

(6) I wanted to gather quantifying data on a seriously large project

I am happy with my progress. [This worksheet](link) [3] ("NewSections") shows that I am ahead of my self-imposed schedule and should have completed recording well before the end of April.

Chapter-size caused me a problem, largely because I was inexperienced.

Of course I had read the wiki and knew that there was a limit of 72 minutes (or 70 minutes?) (or 60 minutes?) so that tracks would fit on a CD, and yes, people do still use CDs, but I was puzzled that LibriVox did not offer a packing tool with its tracks, and that if they did, it would make little sense. LibriVox would need to provide a dis-assembly tool for its readers to make any sense. (But see [this post](link) [4] and [this post](link) [5])

When I receive a set of old VHS tapes to be copied to DVDs, I have a little "best fit" program that takes in the lengths of the movies and clips on the tapes (one or more clips per tape), and takes in the available lengths of space on my DVD collection, and works out the optimum way to cram as much "movie" onto each DVD. I use the same technique to cram as many MP3 tracks from my music collection onto CDs when I am off on a cross-country drive.

But those doesn't work for LibriVox because the 70 Chapters of JSMill have to appear in sequence, so the hapless listener is faced with a sequence (not a series) of audio tracks of apparently random length, only one of which can fit onto each CD. (The 47 tracks I have recorded to date should fit on 22 CDs by my reckoning, but you will need to buy 47 CDs. By the end of the project 94 CDs. What a load!)

Well, so there I was, around Chapter 14 of the project with tracks of 82, 104, 73, 89 and 97 minutes under my belt. And "Book III Chapter V Of The Law Of Universal Causation" threatening to top out at 166 minutes [6].

The LibriVox standards team would insist that five existing tracks be split to make ten tracks, and the Chapter/Section numbers re-assigned. From now on we would drop chatter about Chapters and talk only of Sections. But then the Chapters would start again at "one" when we reached Book 2 of the project, and there are six Books in this work by JSMill.

On top of that there is GREEK. I don't pronounce Greek so we were looking around for a Greek speaker, and in the meantime I did not want to postpone reading those sections with Greek, because I was enjoying reading JSMill and did not want to miss an episode and then come back to it once we had a Greek speaker.

A suggestion was made to split the chapters bearing Greeks into separate bits, with "Greek sections" interlaced. Just like the music sections in the JSBach group project [7] I tarried in JSBach for a while, so I had a view of that tecnique from the sidelines.

I dashed off a quick Word VBA macro that could detect every Unicode Greek character in the HTML file and format it in a character style "csGreek" so that it stood out in my MSWord document.

Of course, once you can find Greek characters, you can identify (and count etc) every word in the document that contains Greek characters. And you can identify each paragraph that contains one or more Greek words. So now the MSWord document looks pretty neat - a direct copy of the official HTML text, but with Greek words in purple, and three font sizes larger than regular text, and the Greek paragraphs identified by a paragraph style "psGreek" that put racing stripes in British Racing Greaves down the left-hand side of each Greek paragraph. Vroom vroom!

Now I can peel out the Greek paragraphs to hand over to a Greek Speaker, if one ever turns up, so that they deal only with a 20-page text rather than 583 pages of abstruse logic.

But now, of course, I can count how many "Greek sections" would need to be spliced in between "English Sections" in any Section (DO pay attention!) of the project, and the number of Sections practically doubled.

Did I mention that:-

(a) I collect billing data [8] in a text file and once we split those early sections I would have to go back and change the task numbers for every Section so far recorded and

(b) I record large sections as parts, so that Section 7, aka "Book I Chapter III Of The Things Denoted By Names; Parts 1 through 5" is assembled from the files S07P01.FLAC, S07P02.FLAC, S07P03.FLAC, S07P04.FLAC, S07P05.FLAC. These files should be renamed; they are used with the Assemble macro to build the master file "logicjohnstuartmill_07_mill_128kb.flac" for the section.

(c) The master files "logicjohnstuartmill_07_mill_128kb.flac" too will need to be renamed, as will the on-the-fly ZIP files I create during recording to safeguard my work.

And look - we haven't started splitting the long tracks yet.

Splitting a track of 104 minutes into two tracks of about 50 minutes each means that every section above that track gets shunted up one number. So Lo! The last shall be first and the sixteenth shall be the twenty-second. But be careful here, the last thing we want to do is drag one set of files over another set; first we must drag the files of Section 22 to the Section 23 folder, then those of 21 to 22, 20 to 21, etc.

Then the Magic Bloody Window had to be extended from 70 to 94 rows (did I mention that this was a learning exercise?) and the titles had to be changed. Where we had had "Book I Chapter III Of The Things Denoted By Names" we now had "Book I Chapter III Of The Things Denoted By Names; Parts 1 through 5" and "Book I Chapter III Of The Things Denoted By Names; Parts 6 through 9".

And no, the MBW will not accept a formatted table, like a CSV file, of data. It must be done by hand.

And the timing figures hh:mm in the notes boxes. And it is Click-on-Save for the Notes boxes but Tap-Enter for the filenames.

Anyway ...

Back into the html and learn what a "silcrow section" is and elect to partition the long chapters into silcrow-sections, except that we can't call them sections because *Sections* is what we have in the MBW, so we had a quick conference, quick because I opted to go with the first suggestion, so now we have Parts.

(There is a good poem there "Today we have naming of parts", unless someone has already grabbed the rights to the title)

Then one cloudy day the sun broke through: Because I already had the HTML text inside the MSWord document, a quick VBA macro could work out where to split each long chapter to the nearest silcrow-section boundary

Odd! It turned out that there weren't as many long chapters as I had thought.

Then I realized that 35,000 words out of JSMill's 474,000 words weren't Footnotes at all, but were Endnotes, hyperlinked to the end of the document, and that is why some long chapters weren't.

Back into VBA and a quick'n'dirty bit of code to drag the footnotes into the text - nothing fancy like plopping them in at the end of a sentence - then back to word counts and silcrow-sections to partition, THEN I could go back and re-title the sections while wallowing in the MBW.

All this cost me about two weeks and I lost the rhythm of JSMill - his manner of writing, his 700-word sentences, his three-line phrases without a commas, and it took me over a week to get back to the point where I could just read the stuff without baulking twice at each jump.

Now, if you dig Latin you will know that a footnote that reads "*Vide supra, book iii., chap. xi.*" or one that reads "*Vide infra, note at the end of § 3, book ii., chap. ii*" Is referring to a chunk of text in a book, and within that book, a chapter. But guess what! Books and Chapters are no good as references in OUR project, especially for blind people, because we now don't really use Chapters. We use Sections.[9]

Have you ever tried on-the-fly translating "*Vide supra, book iii., chap. xi.*" Into "*See Section forty-four of this work, Book three chapter eleven*" or "*Infra, book viii., ch. iv., § 3, and elsewhere*" into "*See Section thirty-eight of this work, Book eight Chapter four Of The Four Methods Of Experimental Inquiry Parts 1 through 4; part four*" without putting the wrong stress in one of those "four"s? I have. It doesn't work very well. I tap "X" to Audacity and pull out the scrap of paper on which I wrote Krista's suggested phrasing, map the book-chapter reference via my worksheet to a section number, write out the new reference, then Record this in the track and exit the footnote hyperlink and get back to the main text.

But anyway, I need not go on; you've got a bit of an idea by now, I hope.

Chris

PS. If you ever decide to do a Solo, take my advice and get a Dedicated Proof Listener almost as good as Krista [10]. I mean, if you can't get Krista because she is flat-out trying to keep up with me, get at least someone who has been *trained* by Krista, OK?

C[11]

[1] I am about to use Footnotes to give you a bit of the taste of the Real World
[2] https://forum.librivox.org/viewtopic.php?f=16&t=91103
[3] http://www.chrisgreaves.com/Downloads/Stats_20220321.xls
[4] https://forum.librivox.org/viewtopic.php?p=1985917#p1985917
[5] https://forum.librivox.org/viewtopic.php?p=1964081#p1964081
[6] See the worksheet "lengths" in the Stats workbook
[7] https://forum.librivox.org/viewtopic.php?f=2&t=66869
[8] https://forum.librivox.org/viewtopic.php?p=1989748#p1989748
[9] I have a sneak-campaign on hand to use the stereo tracks to record page numbers. I mean, if no-one is using CDs then everyone is using stereo players, right? Why not stop discriminating against blind folks?
[10] https://forum.librivox.org/memberlist.php?mode=viewprofile&u=134296
[11] This is the last footnote. Give thanks that there are not 290 footnotes.