20:02 Let's Build a Geniza from the world's Flash Memory!

by Manul Laphroaig

Grace and peace to you!

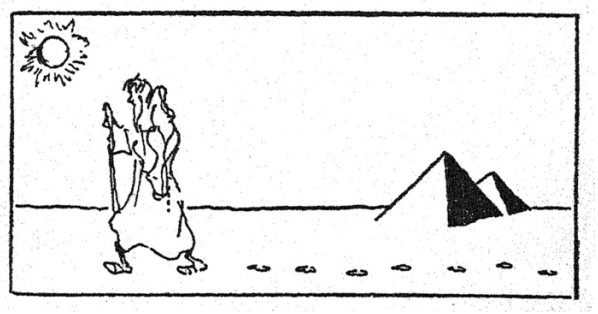
Just this afternoon I finished reading a hundred year old paperback of Thaïs by Anatole France, which thanks to twentieth century mass production cost me as little as I pay for a beer. As I began to marvel that paperback manufacturing has left so many brilliant works of literature in abundance, I also worried for a moment that the ephemeral electronic books of our modern age might leave nothing for future generations. When literature is no longer left around as litter, will my grandchildren be able to afford paper books? Will their grandchildren be able to read?

– — — – — — — – — — –

You see, there was once a fine congregation at the Ezra synagogue in Cairo who believed—as we do—that the written word was sacred. Being at least a little sacred, it wouldn't be right to simply toss their worn out books in the garbage, so the style at the time was to store used and worn out papers in a **גניזה**, a geniza.

They began to store documents in this room nearly twelve hundred years ago, and while every seven years or so they might remove some of these papers for a respectful burial, there were by the end of the nineteenth century some three hundred thousand scraps of writing as a testament to the holiness of inefficient housekeeping.

So the story would have ended, and so similar stories surely have ended in many places and many times in history, except that a professor by the name of Solomon Schechter was given a tattered scrap from this collection. He recognized it as a piece from the Hebrew original of Ecclesiastes, and later recovered the bulk of the collection for indexing and study.



And what might we do, to protect our own books for the long haul? Twelve hundred years from now, as the next civilization is finally printing books and designing computers again after a long, cold night of illiteracy, what treasure trove might we leave for them to print?

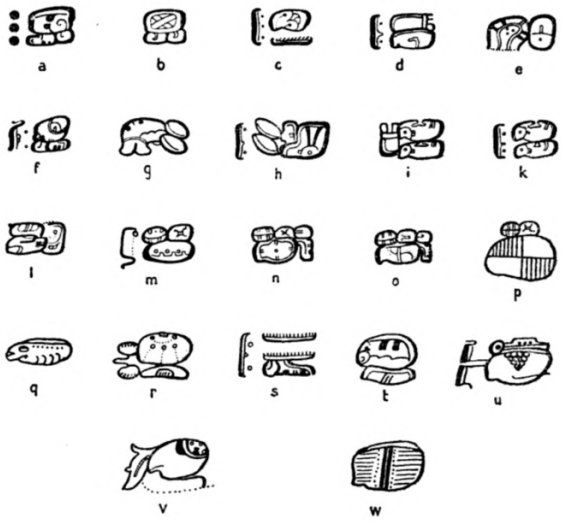
– — — – — — — – — — –

And while I don't mean to be a pessimist, and I don't mean to tell you that the end is nigh, it is a sad fact that civilizations *do* end. I would very much like to see a bit of ours live on.

You see, the written word has been invented three times in history, so far as we know: once in Mesopotamia, once in China, and once in Mesoamerica.

From this third invention, where once there were thousands of books in the Mayan language, just four survived. *Four books* from an entire civilization, all the rest having vanished to the bonfires of a sixteenth century bishop named Diego de Landa.

De Landa, by the way, is not merely one of history's greatest book burners. His own book, Relación de las Cosas de Yucatán, contains the only surviving documentation of the Mayan alphabet, made with little understanding—but with the help of two native speakers. Hundreds of years after his death, this was instrumental to allowing us to finally read the four books that he failed to burn.



And a thousand years from now, what will be found from our civilization, that ancient land in which every man, woman and child carried a black mirror filled with electronics that no longer function? Well, maybe more than we think.

Maybe, just maybe, the next civilization will develop their own computers. Slow ones at first, so let's model them on an Apple][. And having these slow machines with eight bit processors and limited memory, they might realize that the memory chips they've mined from landfills have degraded, but are often still functional.

For a specific example, a SPI Flash chip from a 2010 desktop computer is only a few megabytes, but if you dropped me on a desert island with the parts from an 1980's Radio Shack, it might not take me too long to beep out the contents on an LED if I remembered, or brute-forced, that the read command was `0x03`.[1] It's not unreasonable that a future tinkerer with an eight bit home computer might figure this out as well.

And having one chip, he might try another. Although chips stored in hot environments will have lost their contents, in colder locations it's perfectly reasonable to expect even consumer microcontrollers to hold their contents for a couple thousand years.[2]

And though the denser storage of disks and memory cards will be harder to recover, owing to their dependencies upon the bits of their own ancient firmware, they might still be legible. Except for this pesky modern tradition of full disk encryption, a blessing for personal privacy and a curse to the archivists of the future.

So let's do this:

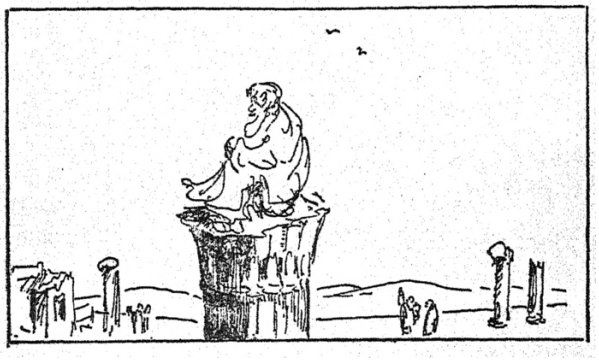
Let's build a geniza of all the text we'd like to preserve, a hundred or so gigabytes worth. All of Wikipedia would consume just tens of gigabytes, and all of Project Gutenberg a little more than six. You can fit this on your laptop.

Let's chop these texts into individually legible fragments, where an encyclopedia article might be ten kilobytes and a novel might be four hundred.[3] We want each fragment to be individually meaningful, and while some chunks will surely be erased and overwritten, those that survive ought to be easy to re-assemble.

Let's write a utility that can summon one or thousands of these fragments on demand, organized into batches of the native block size. A bit of light compression or error correction won't hurt, but like error correction in the POCSAG standard, this one should be optional and off to the side, so as not to hide the meaning of the message.[4] Where the device has full disk encryption, this must be outside of the encrypted region, but it is perfectly okay that many of these blocks will be destroyed as the operating system claims those blocks for its own use.

And finally, let's use this tool to stuff *every unused block* of memory with literature at the factory! Whether the ten kilobytes that will never be used in my wristwatch or the hundred gigabytes not yet used in a cellphone, let's fill *all* of the spare space in these chips with a geniza for the future.

Done right, in the test routines of a major product, one single engineer might seed every landfill in the world with these books, not just in a single generation, but in a single year! And if you are that engineer, I will very happily buy you a beer.



[1] `unzip pocorgtfo20.pdf w25q128fv.pdf`
[2] `unzip pocorgtfo20.pdf flashretention.pdf`
[3] `unzip pocorgtfo20.pdf 80days.txt revolt_en.txt thais.txt`
[4] `unzip pocorgtfo20.pdf pocsag.pdf`