

2015

Pelikan's Antidisambiguation: Editions, Tweaks, and User Preferences

Michael P. Pelikan

Penn State, mpp10@psu.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Pelikan, Michael P. (2017) "Pelikan's Antidisambiguation: Editions, Tweaks, and User Preferences," *Against the Grain*: Vol. 27: Iss. 2, Article 44.

DOI: <https://doi.org/10.7771/2380-176X.7068>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Little Red Herrings
from page 73

in a print world, enjoyed Sardanapalian benefits, are trying to recapture those cash cows in bits and bytes but with little success. It isn't so easy, but they're discovering it is much cheaper to print an electronic book while dropping the price only marginally. Like online courses at war with classroom ones, online books are going to be cheaper and provide a greater return on investment. That ROI does not necessarily include what students are investing in, however. If eBook reading increased 200%, it would still have a way to go before it caught up with print reading if measured in terms of value received and retained.

What this means for libraries is obvious, isn't it? We still have to collect and support both for the time being, in the same way that we have for years supported microfilm and bound periodical volumes. Microform reading only caught on when there was no other choice. I would find it surprising if eBooks end up in the same dustbin. Microform-reading was never easier, better, or more convenient. Nothing about it enticed the reader. Its only attraction was a pedestrian one: it saved space while still providing access, even if a difficult one. eBooks have already shown their value in the benefits mentioned above, but also in leisure reading. None of us really like lugging suitcases of print books with us on vacation (my long-suffering wife will argue that she knows at least one person). Having the ability to take literally hundreds appeals to those of us with eyes larger than our brains.

But when it comes to scholarship that must be recalled and remembered, few of us will choose the electronic text over its printed counterpart. I believe this to be more a facility of evolution and practice rather than something inherently hard-wired in us. Unless or until we can rewire our brains — and, for better or for worse, online reading is doing that — we will read both formats, depending on the subject matter and/or reason for reading.

I haven't had time to sift through the new literacy report, so I cannot speak to how well or to what extent the issue of online reading contributes to the strength or weakness of it. If the students in the **Rosenwald** story are right, and if my own research in this subject matter is at all correct, it may well unravel many of the gains we have made in recent decades. Poor readers, especially, will have a much tougher time going forward if they must learn to read digitally first. If that continues, we will see future generations underperforming when compared with their past peers.

PRENAX
an International Company

BASCH
a Prenax Company

Subscription Management Solutions for Libraries & Corporate Procurement

- ▶ Subscription management
- ▶ E-procurement integration
- ▶ E-journal set up and activation
- ▶ E-journal URL maintenance
- ▶ Click-through access to e-content
- ▶ Cost center accounting
- ▶ Automatic claiming
- ▶ Custom and branded e-portals
- ▶ License negotiation and management
- ▶ Flexible management reporting
- ▶ Built in approval process
- ▶ Express payments to publishers
- ▶ Check in option for print titles
- ▶ Partnerships that provide usage statistics, rights management, discovery tools and single sign on.

Prenax Inc. provides subscription management solutions for procurement professionals and libraries. As a partner, we provide a single point of contact for managing electronic and paper subscriptions, professional memberships and books. We offer a true one-stop shop for all business, scientific, technical, medical, research publications and electronic content. We save you time and money and eliminate the hassle of working with multiple content suppliers.

Prenax offers the flexibility of two platforms, one for servicing libraries and one suited for serving corporate customers.

Basch Subscriptions, Inc.
Prenax Inc.
10 Ferry Street, Suite 429, Concord, NH 03301
(P) 603-229-0662 (F) 603-226-9443
www.basch.com • www.prenax.com

And so, the print versus online debate continues in its ironies, even as you read this article first in print, or, if you come to it much later online. 🍷

Pelikan's Antidisambiguation — Editions, Tweaks, and User Preferences

Column Editor: **Michael P. Pelikan** (Penn State) <mpp10@psu.edu>

I've made comments before in this space about problems that continue to plague eBook projects that begin with out-of-copyright print sources. Optical Character Recognition (OCR) has improved hugely over the past ten or fifteen years, but achieving the last incremental improvements that would bring it close to practical perfection has proven difficult. Even if achievable, near-perfect OCR would do nothing to address the backlog we've accumulated of poor OCR'd texts, many of which, as mentioned, are out of copyright.

This means there's not a lot of financial incentive to promote investment in retrospectively repairing past results of flawed OCR projects. This came up for me again recently whilst reading, for only the second time in my life, the *Personal Memoirs of Ulysses S. Grant*.

My first encounter with this material was through Project Gutenberg. It came in the form of a pure ASCII text file. It had line endings and carriage returns, but nothing more exotic than that. The file itself was not the product of OCR. Instead, it was typed by true enthusiasts: candidates for sainthood who felt strongly enough about a particular book to take on the task of transcribing as an entire work from printed page into keystrokes, for the good of the World.

The quality of transcription of many such works was variable, but improved over time. This was not in small measure because other folks came along and began to make corrections to the hand-built editions, in a way somewhat similar to how a wiki article can be improved over time. Better, in some ways, because there were fewer matters relying upon

subjective interpretation, at least in the case of same-language transcriptions — either it was correct or not.

I don't really understand, if a human-generated, even curated, transcription exists, why the builders and publishers of e-texts don't take advantage of them. Why start from scratch and apply machine-driven OCR to printed text if there's already a transcription? Many, perhaps most, such transcriptions are freely available and could be used — it would cost only attribution and recognition of the source, something I'd perhaps wrongly assume that even the most craven, financially motivated republishers of old works could bring themselves to do.

Instead, now, a dozen or more years after admiring the transcription of **General Grant's** memoirs, and hoisting a coffee cup in toast to

continued on page 77

Pelikan's Antidisambiguation from page 75

the unknown person or persons who made it possible for me to enjoy the work, I'm confronted with obvious, characteristic OCR errors in a recent eBook edition. Grumble.

But this shouldn't be the end of the story! Have you noticed that **Kovid Goyal's Calibre** (<http://caliber-ebook.com>) permits the editing of an eBook file? Regular readers of "**Anti-disambiguation**" (at least, those who would admit to it) will recognize my shout-out to this extraordinary open source software package. If you use an eBook reader, I mean, at all, you owe it to yourself to have a copy of **Calibre** installed somewhere.

All right, but say I use **Calibre** to fix an obvious OCR botch in an out-of-copyright work like **Grant's** — what then? Well, I'd have to sync the repaired file to the several eBook readers I maintain, as well as the file servers I keep at home for purposes of redundant backup. Ever looked into NAS RAID devices? These are a faintly miraculous technology, once accessible only among the corporate or the hopelessly geeky — now available to all! I presently employ three of these boxes on my home network, each containing two hard disk drives configured to mirror each other. Whilst they quiet the mind, they also exact a bit of overhead in terms of file management — but good file management will always entail a blend of good decisions and good practices.

The idea of applying corrective measures to an eBook differs only in degree from things we already do. Those controls on your audio devices labeled Bass and Treble? Those have been collectively referred to in the past not merely as Tone controls, but Equalization controls. The concept behind audio equalization is corrective. Recognizing that different listening environments have differing acoustic characteristics, as do the many and various transducers in use, thoughtful manufacturers of audio gear provided audio controls permitting one to tailor the frequency response of one's

audio gear to compensate. If your rugs and curtains absorb high frequencies resulting in, say, a six dB roll-off at 10 kilohertz, you can boost the response of your system at 10 kilohertz by six dB to "equalize" it.

Of course, many folks don't use these controls to equalize anything but, in fact, to de-equalize, indeed, to change the frequency response of their audio systems simply to suit their preferences. Those worthies cruising slowly down the street in the low car with dark windows and after-market muffler, whose audio system's subwoofer can be heard two blocks away, sending ripples through puddles like **Crichton's** T-Rex, melting their tympanic membranes — they're merely applying user preferences.

This appetite to configure, to tweak, to personalize, must cause despair, or at least shrugs, among the engineers and producers who struggle to achieve a particular sound in a produced recording. The thoughtless destruction of producer's and artist's wishes has been going on for a long time. Ever been in a discount store and heard one channel of a stereo recording in housewares and the other in lawn and garden? I recall a story my brother told of the fourth and last time he went to **Stanley Kubrick's** "*2001, A Space Odyssey*" — it was in 1969 at a drive-in theater in Indiana. It was raining heavily. You could just make out the screen through the fogged windows. The little metal speaker box hanging in one side window was struggling to handle "Also Sprach Zarathustra" with little success. Poor little thing...

I've long wished for there to be released the audio version of critical editions of recorded classics. As a darn-near-life-long multitrack audio production guy, there's nothing I'd like more than to get my hands on a multitrack version of particular classic recordings. As soon as the **Beatles** got past "Beatles '65" they were increasingly taking advantage of technical possibilities afforded them by their studio,

and opened by the skills of **George Martin**. **Hendrix's** early recordings were very simple. In the space of a few hundred days these artists were taking their music places few had gone before, and they were layering sound upon sound to do so. It was the audio equivalent of photo or motion picture compositing, placing elements of differing origin into seamless proximity with each other.

With a multitrack edition of these recordings, one could separate the original signals, listen to each individually, and gain a better understanding and appreciation for how the producer and the artist achieved such phenomenal results. Of course, it would require that a multi-channel mixer be part of the signal chain — but who wouldn't want that? And if a particular sound always seemed buried to you, you could bring it out in the mix! Conductors do this when they interpret a score in front of them, shaping the statement and balance of each of the parts of the score through guidance provided to the orchestra. Really, a musical score is a multitrack representation. So its counterpart in recorded music — that's all I'm asking for...

Blu-Ray and DVD editions of motion pictures often offer options in playback to include or exclude deleted scenes, to change language settings, etc. I've seen the occasional book, usually a children's book, that feature branching in the storyline, permitting exploration of alternate plotlines based upon decisions as you go.

I know it will probably not happen in my life time. Works of interpretation are works themselves — that's probably part of the reason why such a great idea won't easily come about. Royalties and Intellectual Property issues involving derivative works get complicated. But I'd be happy to sign a license attesting that I would not release a remix of Sergeant Pepper or Electric Ladyland — I would only take bits of them apart to see how they work. This isn't too different from standing in front of an artist's masterpiece in a museum with a sketchpad, working with charcoal and paper to understand what's going on in the painting or sculpture.

There are some promising prospects enabled by digital audio analysis. Some of the same algorithms that achieve noise removal through example (sample the offending waveform, then look for it in compound waveforms and separate it out, leaving a clean signal) can be used to "de-mix" a mixdown. It might be feasible before long to divide a favorite recording back into separate tracks.

If you're interested, there's an intriguing PhD dissertation at **Stanford's** Center for Computer Research in Music and Acoustics entitled "Interactive Sound Source Separation" by **Nicholas J. Bryan**. The dissertation is licensed under a Creative Commons Attribution-Noncommercial 3.0 United States License. Google that title to find the pdf. Outstanding work. 🌸



Let's Get Technical from page 76

triage, as well as to isolate any books that have mold.

After performing the check of bibliographic elements and changing the location code, books are reshelfed in the staging area on their spine. This signals to student workers that these books are ready for boxing to be sent to the off-site storage facility. Students track the number of books boxed on a spreadsheet so that we can report to our administrative office the number of volumes being sent, since we are charged by our consortia's main office for relocating them to the off-site storage facility. Our consortia office sends a truck to pick up the boxed books once a month.

The Results

Since the move to storage project began in June 2012, we have relocated approximately 102,000 volumes to off-site storage, freeing up student study space within the library. The secondary benefit to the project has been cleanup of records and ensuring that books match the record they are attached to before sending them to off-site storage. This is in addition to evaluating the books conditions and having the opportunity to repair or replace damaged or moldy materials. 🌸