

Digitizing Sample for „The Yoga of Pantanjali“

A digitizing test from Benjamin - 21.Feb 2021. Questions please to benjamin.schiller.mail@gmail.com

The idea of this test was to find a good digitizing process, so that other friends can do the same without the hurdles while doing the test.

Book preparation

For the book preparation you need:

- a book in a good shape - knowing the book is after that process in a bad shape!
- a sharp knife - use it carefully and slowly
- a paper cutting machine like below on the right



1. Scan cover/back as jpeg in good quality **before starting with cut off**
2. Cut off the cover
3. Cut off the connecting edge of the pages

The original book came in a bad quality from Ludger, but it was the last exemplar he could find. So better than none and a perfect test object for getting a good quality also with heavily used material:

- Book corners strongly bumped
- yellowed paper
- partly strongly shifted page beginnings like handmade copies
- the first and last pages strongly glued to the cover - hand made

Scanning

There is a lot of scanning software. Personally I use XSane (Linux) for automated scanning. Not the best user interface but clever features. So I recommend XSane.

If you are using alternative tools on other operating systems, they should support

- suffixing of scan files and
- good gamma and contrast correction
- and additional if possible: straightening, cleaning and all the other options which are helpful for the post processing

Digitizing Sample for „The Yoga of Pantanjali“

A digitizing test from Benjamin - 21.Feb 2021. Questions please to benjamin.schiller.mail@gmail.com

For scanning (with XSane) you need:

- Scanner with paper feeder. Manual page scan is not recommended.
- In my case Linux client e.g. Ubuntu > 16.04, > 2 GB RAM or your computer / OS which fully supports your scanner

1. Optimized scan with automatic sheet feeding e.g. with XSane

- automatic page naming
- good contrast (+40) correction for a clean grayscale (black/white) scan
- scan the odd pages with 300dpi beginning with the first page *<book name>-0001* and a +2 pages suffix after each scan. *The idea is to scan the book as a stack of paper.*
- scan the even pages with 300dpi beginning with the last page *<book name>-0178* and a -2 pages suffix after each scan. *The idea is simply to turn over the stack of paper.*

2. Add the cover with page 0000 and the back with *last page+1*

Post processing

Post processing is the part for automated straightening and cleaning of scan files. ScanTailor is an easy and powerful tool for doing that on GNU/Linux or Windows. There could exist alternative tools on other operating systems. They should support:

- "auto straighten" of skewed scans
- "auto content select" for removing scan shadows at scanned page borders

I do not recommend manually cleaning page by page.

1. With ScanTailor you have the following options:

- input format jpeg and a few others
- very important: "auto straighten" of skewed scans
- very important "auto content select" for removing scan shadows at scanned page borders
- automated cleaning of the scan like: orientation fix, page split, margins, resolution, ...

2. ScanTailor uses jpeg format and a few others produces cleaned tiff formatted files, an old format but convertible.

- Convert tiffs to one pdf: *convert *.tif -compress zip -quality 100 out.pdf*

3. [Compress the one PDF to a book in ebook quality \(resolution\)](#) - a really good! compression. In my case 114 MB down to 14,6 MB. And the result is easy to read.

- `gs -sDEVICE=pdfwrite -dCompatibilityLevel=1.4 -dPDFSETTINGS=/ebook -dNOPAUSE -dQUIET -dBATCH -sOutputFile=the-yoga-of-patanjali.pdf out.pdf`