See Also:

Digital Aladore: Reading and Creating with Free Software and the Public Domain

Evan Peter Williamson - evanpeterw@gmail.com

Dual MAS/MLIS

Abstract:

I love reading old books and I love reading on my e-reader. I also love free stuff—cost free and freedom-free, like Free software. Well-known authors in the public domain have many high quality ebook editions freely available from organizations such as Project Gutenberg or commercial vendors such as Feedbooks. But what if you want to read something more obscure?

I was trying to read Henry Newbolt's Aladore, a "forgotten fantasy" novel from 1914. Print editions were digitized by the Internet Archive, but a true ebook version was never created. The digitized books can be read in an online viewer or downloaded as an image-based PDF. Unfortunately, the PDFs are too highly compressed, requiring extensive rendering time in any reader. Internet Archive also provides numerous other formats derived from automated OCR. Unfortunately, the quality is very poor and there has been no attempt to edit the resulting text or format the ebooks. Basically, you have a choice between a low-quality-ridiculously-slow-and-cumbersome PDF or a gibberish-filled-automatically-generated EPUB. This makes for a horrible reading experience!

It was so horrible, I decided to create my own ebook. Thus, Digital Aladore (https://digitalaladore.wordpress.com) was launched September 2014—exactly one hundred years after Aladore was first published.

The idea is to use freely available public domain materials (the digitized copies of Aladore) and free software to create a GOOD digital reading edition of the text, and to blog about the entire process along the way. The bigger idea is to demystify the creation of ebooks, empowering readers to be reflective creators. Digital Aladore has a zero dollar budget: public domain content, free software, recycled hardware—all you need is some interest, passion, and perseverance. The blog explores preserving, creating, and sharing through public domain, free software, open formats, and Creative

Commons. It reflects about the process of digitization and textual transmission. However, the main focus is a practical hands-on spirit: crack open an EPUB or an old computer, and look inside!

The project provided a great opportunity to explore, learn, and practice a huge variety of things that I only vaguely knew about at the beginning. Attempting to carry out tasks—then encountering issues—then searching for solutions—then articulating the ideas in the blog was a great way to build concrete skills and give body to theoretical concepts. Blogging was also a great experience where I was able to test different ideas about communication and access, while connecting with an interesting mix of other bloggers, some of who became important resources for the project.

Digital Aladore has produced several ebooks, public domain images, text visualization demonstrations, and detailed documentation for an ebook creation workflow using numerous software applications. The poster outlines the main ebook creation process and how it relates to the public domain, free software, and Creative Commons.

Copyright:

Digital Aladore:
Creating and Reading with Free Software and the Public Domain.

Henry Newbolt's *Aladore* is an obscure fantasy novel from 1914. It was digitized by Internet Archive, but a true ebook version was never created. So you have a choice: do you want to read the low-quality-ridiculously-slow-and-cumbersome PDF or the gibberish-filled-automatically-generated EPUB?

I started the Digital Aladore project to create a DYI alternative, a high quality ebook edition on zero budget to be freely distributed. To share what I learned about the process, I blog about all the details—because I want you to go out and make some ebooks for me to read!
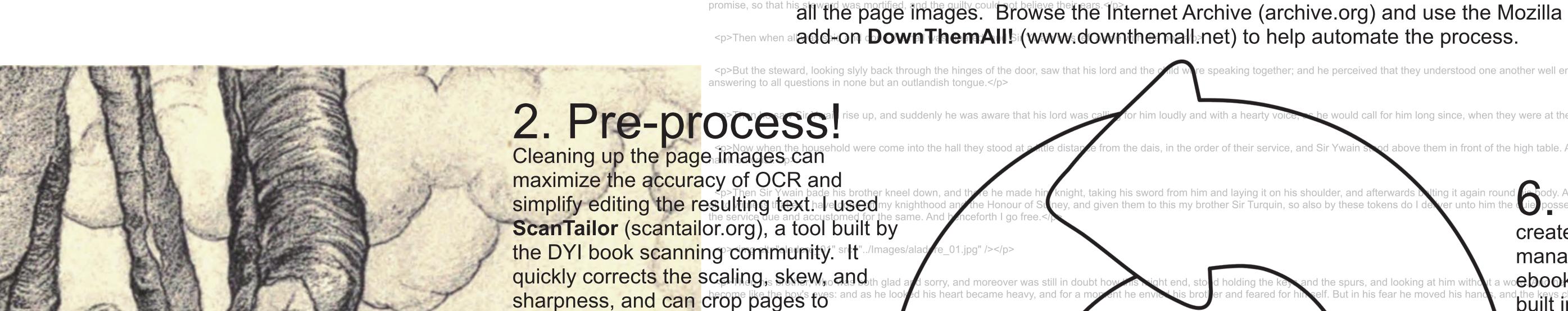
What is Free Software?

"... software that respects users' freedom and community. Roughly, it means that the users have the freedom to run, copy, distribute, study, change and improve the software."
(https://www.gnu.org/philosophy/free-sw.html)

All the applications used in Digital Aladore are Free software, both cost-free and free as in "free speech" and liberty. This means the source code is distributed and licensed so that anyone can use, modify, and share it. Based on a philosophy of transparency and collaboration, free software allows everyone to build on each other's work and put control in the hands of the user. It was vital for this project to find tools that fully empower people to create and share.

1. Download!
Find a digitized public domain book and download all the page images. Browse the Internet Archive (archive.org) and use the Mozilla add-on **DownThemAll!** (www.downthemall.net) to help automate the process.

2. Pre-process!
Cleaning up the page images can maximize the accuracy of OCR and simplify editing the resulting text. I used **ScanTailor** (scantailor.org), a tool built by the DYI book scanning community. It quickly corrects the scaling, skew, and sharpness, and can crop pages to eliminate the header and footer.

3. OCR!
Optical Character Recognition lets your computer read, converting the page image into machine-readable text. I used the **Tesseract** OCR engine (code.google.com/p/tesseract-ocr), an open source tool supported by Google. It can be used via command line, or with numerous GUI applications. I used **YAGF** (sourceforge.net/projects/yagf-ocr) and generated an HTML transcript from the page images.

4. Create Ebook!
The raw OCR transcript has many small errors and gibberish characters. A lot of editing is required to create a Good reading text. I used a number of tools to help simplify the process, from the handy **BlueFish** text editor (bluefish.openoffice.nl) to the scholarly collation tool **Juxta** (www.juxtasoftware.org). Once the text is ready, I create an ebook using the full featured EPUB editor **Sigil** (github.com/user-none/Sigil).

OPEN!
Public Domain Book
+ CC-by-sa Blog content
+ FOSS applications
+ Open formats
= Public Domain eBook

5. Distribute!
The final ebook and texts are freely released on Internet Archive in a variety of formats for use and re-use. To make sure the practical know-how is also distributed, the Digital Aladore blog documents every detail of the process, sharing the experiences, tools, and workflows. Close the loop by sharing alike!

6. Read!
Download some free ebooks or create your own, and read! If you need a great ebook management application, check out **Calibre** (calibre-ebook.com). It has a reader, format converter, and editor built in: a powerful tool to explore your ebooks inside and out! Get the draft Aladore EPUB at: https://archive.org/details/AladoreNewbolt

https://digitalaladore.wordpress.com
Demystifying the creation of ebooks, empowering readers to be reflective creators.
Public domain content, Free software, Recycled hardware:
all you need is some interest, passion, and perseverance!

Evan Peter Williamson
evanpeterw@gmail.com
University of British Columbia, 2015
iSchool@UBC
SCHOOL OF LIBRARY, ARCHIVAL & INFORMATION STUDIES