BVCC General Meeting

April 14, 2025

PDF Utilities and Tools

Joel Ewing, BVCC President

These slides will be published on the BVCC website
(Information ►Presentations)

PDF File Format

- **Portable Document Format**
  - **Designed by Adobe to accurately represent the visual appearance of a printed document in digital form as a combination of text (with font info), images, and vectors (scalable lines) and related positional information**
  - **A ubiquitous format that can be universally displayed**
- **Many word processing applications can save a document as a PDF – but,think of it as a digital printed document, not something designed with editing in mind.**
  - **Each line of text is saved as a separate text block, each image as an image block – can search for words. provided they are in text block, not embedded in an image.**
  - **Contextual info is lost, only knows text lines, images and their position on the page, not document structure.**
  - **Accurate reverse translation back to a word processing format is in general not possible.**

PDF File Format

- **Image editors and scanners can also produce image-format PDF documents**
  - **Each PDF page is just one big image**
  - **Text that is in an image cannot be searched**
  - **If loaded into most word processors, will become an image in the document with no ability to edit embedded text.**
- **Some software may support OCR to add actual text fields to a PDF**
  - **Will attempt text recognition and add text blocks with recognized text to the PDF underneath the image at the approximate location for search**
  - **Accuracy of OCR heavily dependent on font quality and image resolution and is never 100% perfect**
  - **Resulting image/text PDF may be searchable for words**
  - **May even be possible to select and copy some lines of the PDF document and paste text content into a word processor document, but expect a lot of manual editing.**

PDF File Format

- **PDF Forms vs PDF Documents**
  - **PDF form is a PDF Document with specific fields that can be entered/edited without a generalized PDF editor utility**
- **PDF protection options**
  - **Can allow read without allowing update**
  - **Can encrypt to prevent any access by any program without a key**

PDF File Format – Vector PDF

- **Some tools can create Vector PDF (images and text using lines of varying width, direction and color, or with color fill of areas bounded by lines)**
  - **Less common than other PDF formats**
  - **Simple images with larger, same-color areas can converted to vectors.**
  - **Faithful rendition at high zoom – lines are re-computed, pixels not magnified.**
  - **Vector representation can be very space efficient, but**
    - **Lose ability to edit or search for text only saved as vectors**
    - **Not suitable for hi-res finely-detailed images or images that require full 24-bit color range.**

Useful Utility Functions for PDFs

- **Merge multiple PDFs to single PDF**
- **Split a single PDF into multiple PDFs**
- **Delete, rotate, shuffle the order of individual PDF pages**
- **Convert multiple image files into multi-page PDF (with OCR)**
- **Convert multi-page PDF into multiple image files**
- **Add text fields to PDF pages – available fonts may be limited**
- **Edit existing text fields in PDF**
  - **Can be tedious – no auto-wrap or justification across text field boundaries**

Examples of PDF Tools

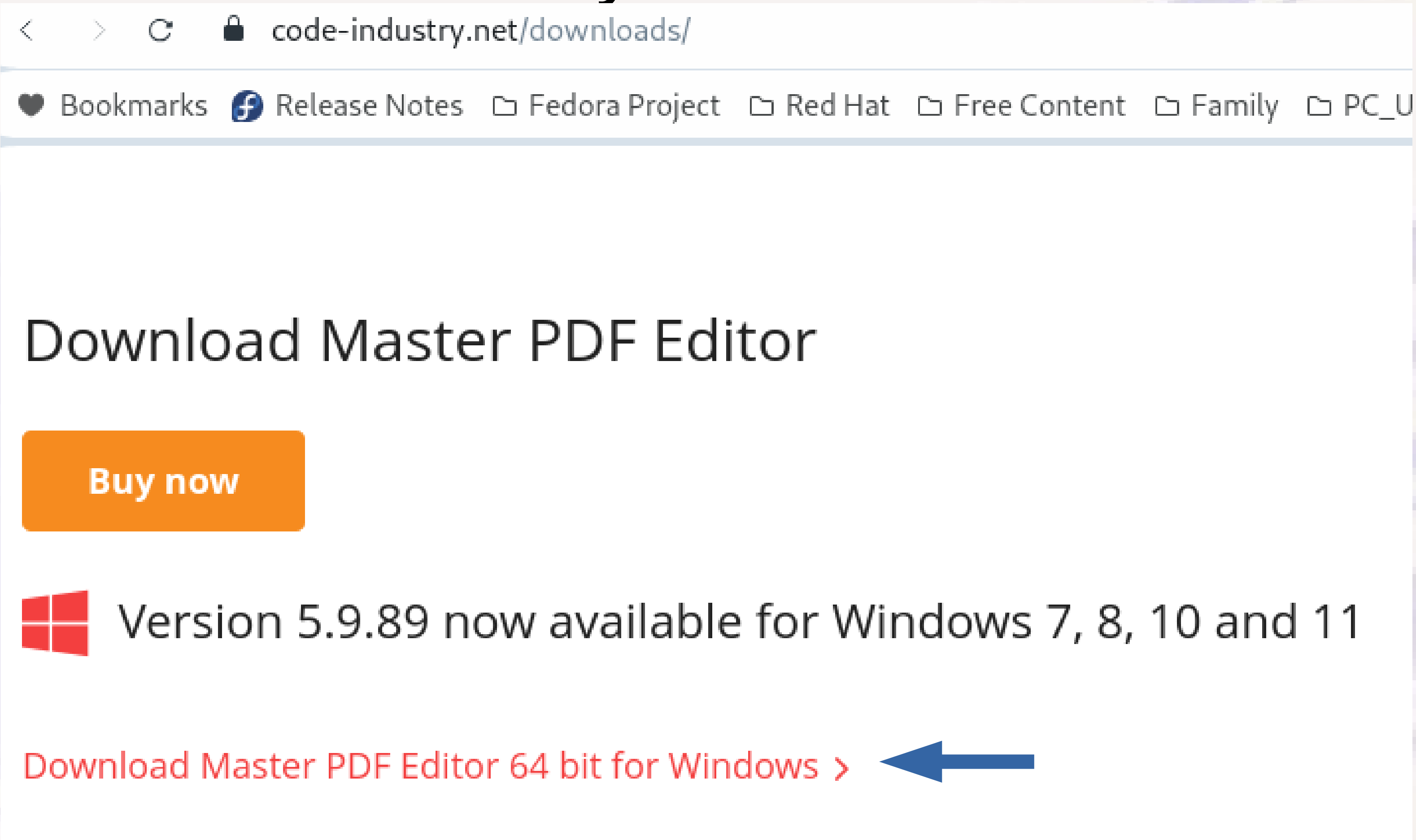
Adobe Acrobat

- **Standard ($155.88/yr), Pro($239.889/yr)**
  - **A full-featured editor**
  - **Hard to justify cost if only an occasional user**
  - **Most users will find lesser editors adequate**

Code Industry Master PDF Editor

- **Multi-platform, Windows, Linux, Mac**
  - **$79.95 one-time purchase https://code-industry.net/downloads/**
  - **FREE & fully functional, if you don't mind a watermark on each page after the 30-day trial period.**
- **Can perform all the basic operations listed earlier, plus some additional ones.**

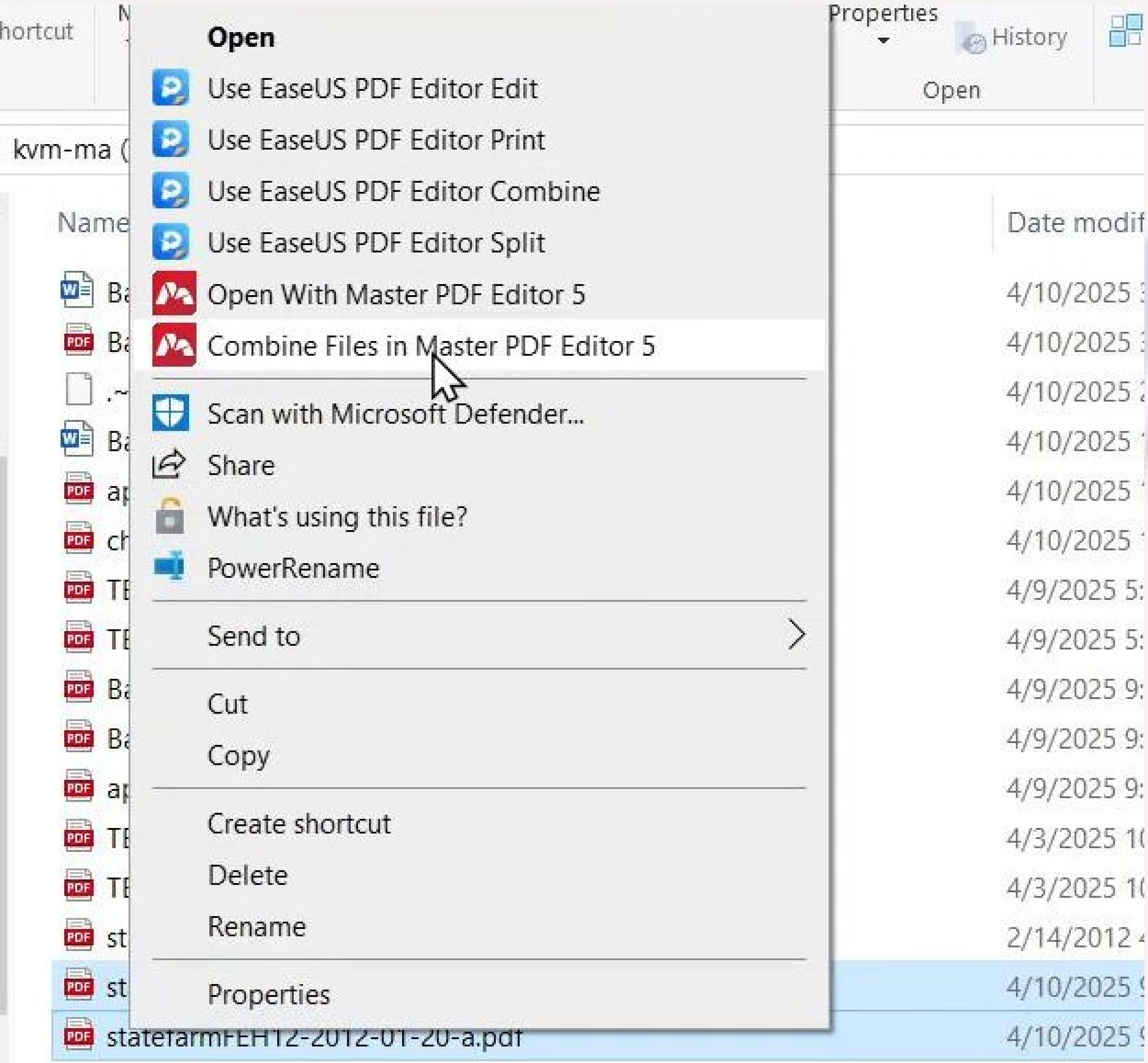
Code Industry Master PDF Editor



Download Master PDF Editor

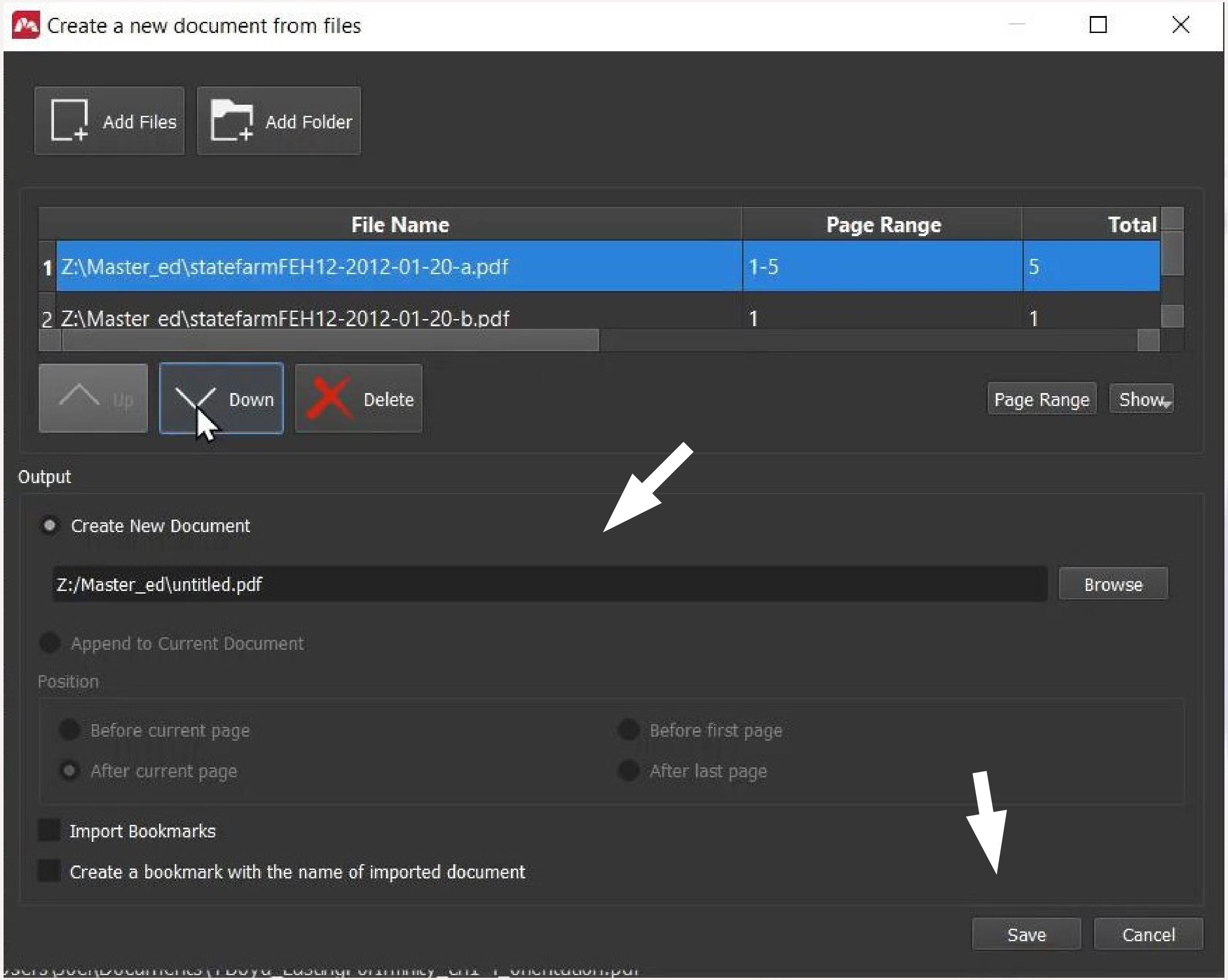
Version 5.9.89 now available for Windows 7, 8, 10 and 11

Download Master PDF Editor 64 bit for Windows ›


hortcut
Properties
History
Open
kvm-ma (
Name
Date modif
Open
Use EaseUS PDF Editor Edit
Use EaseUS PDF Editor Print
Use EaseUS PDF Editor Combine
Use EaseUS PDF Editor Split
Open With Master PDF Editor 5
Combine Files in Master PDF Editor 5
Scan with Microsoft Defender...
Share
What's using this file?
PowerRename
Send to
Cut
Copy
Create shortcut
Delete
Rename
Properties
statefarmFEH12-2012-01-20-a.pdf
4/10/2025
4/9/2025
4/3/2025
2/14/2012

Combining PDFs

Select multiple PDF files in File Manager, right-click and select "Combine Files in Master PDF Editor"



	File Name	Page Range	Total
1	Z:\Master_ed\statefarmFEH12-2012-01-20-a.pdf	1-5	5
2	Z:\Master_ed\statefarmFEH12-2012-01-20-b.pdf	1	1

Output

● Create New Document

Z:/Master_ed\untitled.pdf

○ Append to Current Document

Position

○ Before current page
● After current page
○ Before first page
○ After last page

□ Import Bookmarks

□ Create a bookmark with the name of imported document

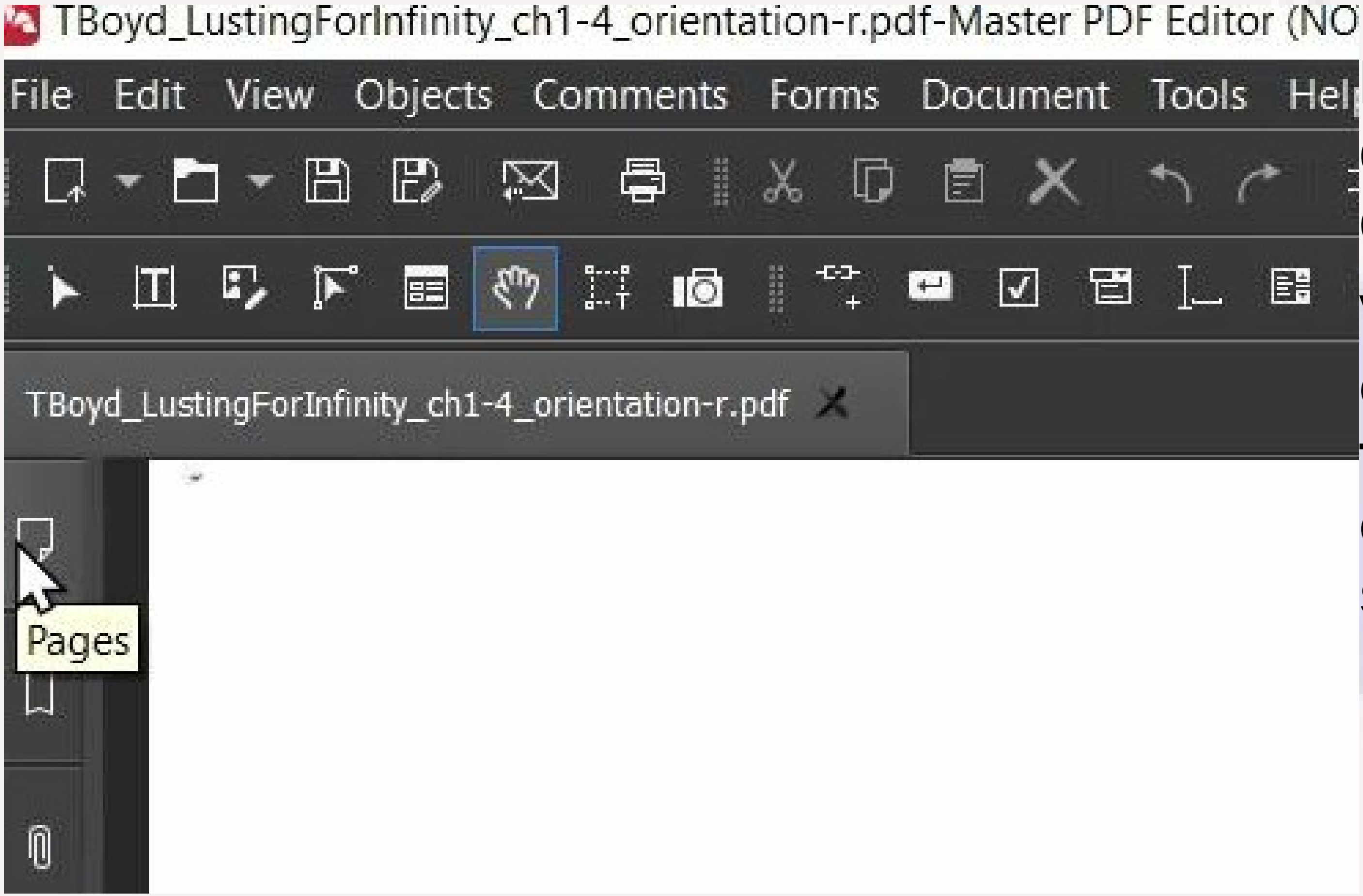
Combining PDFs (cont)

If files are not in correct order, select a file and move-up or move-down, etc.

When files are in desired order, enter correct output file name. Can "Browse" to correct directory and then modify file name as needed. Then "Save" to store the combined PDF.

Default action after saving is to open the new document in the Master PDF Editor

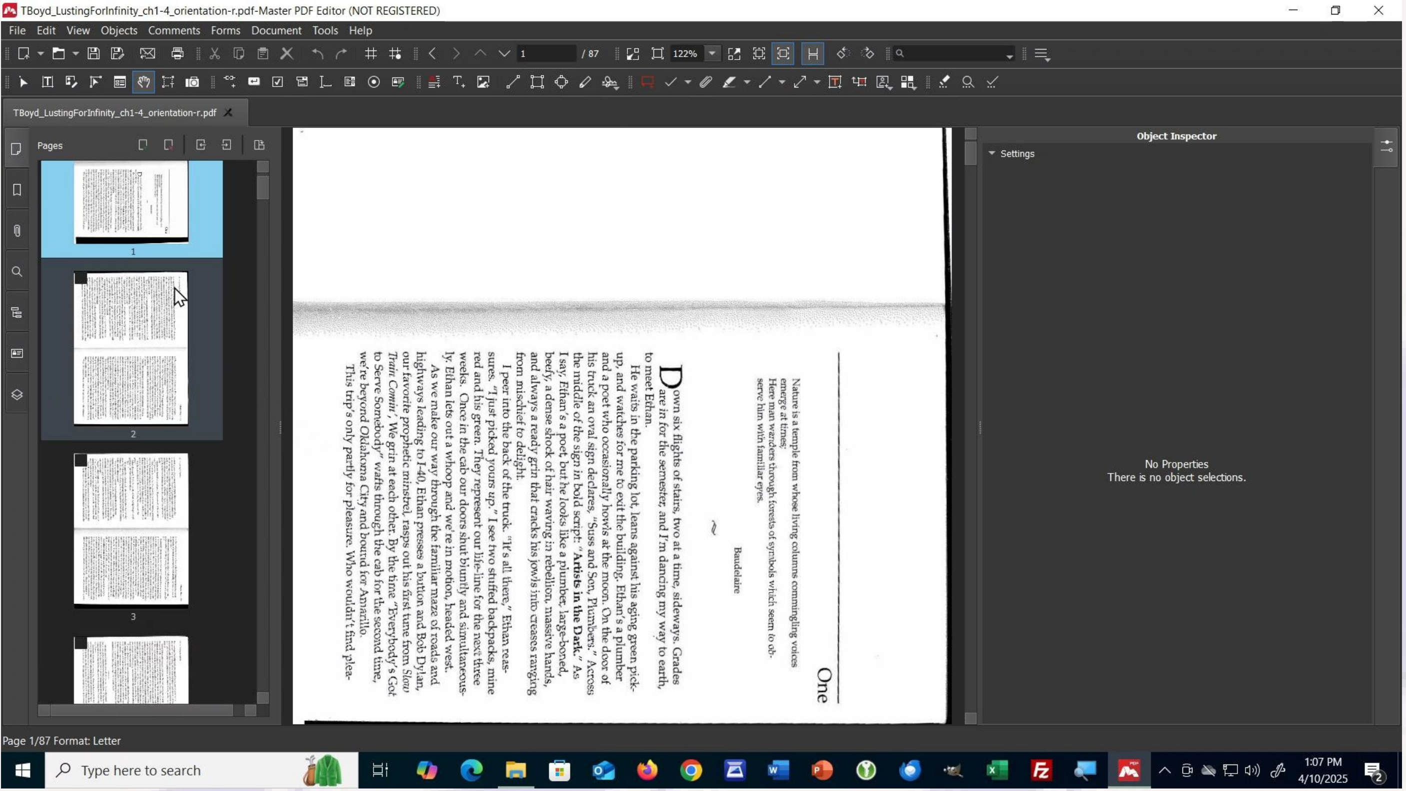
Editing (open with Editor)


TBoyd_LustingForInfinity_ch1-4_orientation-r.pdf-Master PDF Editor (NO
File
Edit
View
Objects
Comments
Forms
Document
Tools
Hel
TBoyd_LustingForInfinity_ch1-4_orientation-r.pdf
Pages

Begins with display of part of document.

Very useful to click on "Pages" to get thumbnail images of all pages for selection purposes

Editing


TBoyd_LustingForInfinity_ch1-4_orientation-r.pdf-Master PDF Editor (NOT REGISTERED)
File
Edit
View
Objects
Comments
Forms
Document
Tools
Help
1
/ 87
122%
TBoyd_LustingForInfinity_ch1-4_orientation-r.pdf
Pages
1
2
3
Object Inspector
Settings
Type here to search
1:07 PM
4/10/2025

One

Nature is a temple from whose living columns commingling voices emerge at times;
Here man wanders through forests of symbols which seem to observe him with familiar eyes.

Baudelaire

Down six flights of stairs, two at a time, sideways. Grades are in for the semester, and I'm dancing my way to earth, to meet Ethan.

He waits in the parking lot, leans against his aging green pick-up, and watches for me to exit the building. Ethan's a plumber and a poet who occasionally howls at the moon. On the door of his truck an oval sign declares, "Suss and Son, Plumbers." Across the middle of the sign in bold script: "**Artists in the Dark.**" As I say, Ethan's a poet, but he looks like a plumber, large-boned, beefy, a dense shock of hair waving in rebellion, massive hands, and always a ready grin that cracks his jowls into creases ranging from mischief to delight.

I peer into the back of the truck. "It's all there," Ethan reassures. "I just picked yours up." I see two stuffed backpacks, mine red and his green. They represent our life-line for the next three weeks. Once in the cab our doors shut bluntly and simultaneously. Ethan lets out a whoop and we're in motion, headed west.

As we make our way through the familiar maze of roads and highways leading to I-40, Ethan presses a button and Bob Dylan, our favorite prophetic minstrel, rasps out his first tune from *Slow Train Comin'*. We grin at each other. By the time "Everybody's Got to Serve Somebody" wafts through the cab for the second time, we're beyond Oklahoma City and bound for Amarillo.

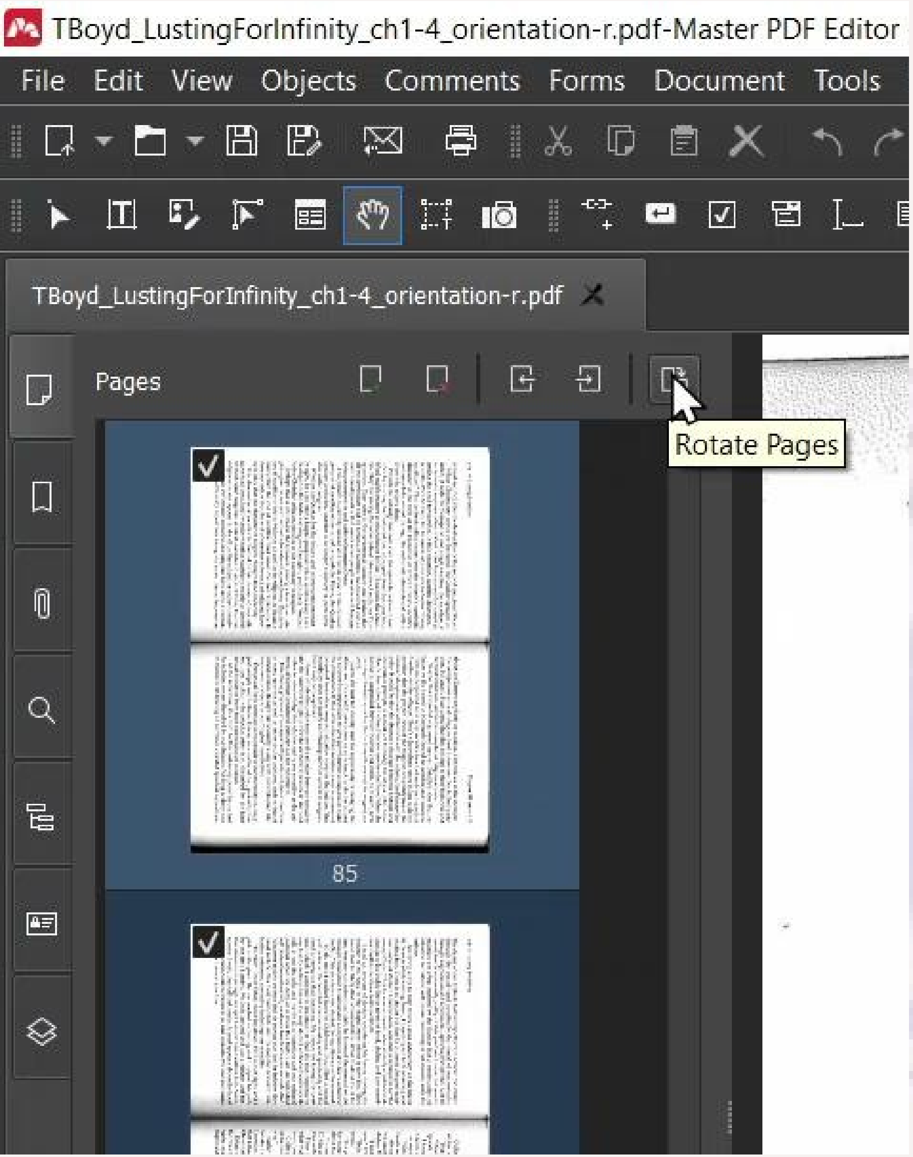
This trip's only partly for pleasure. Who wouldn't find plea-

No Properties
There is no object selections.

Page 1/87 Format: Letter

Thumbnail images on left can be used to select and quickly move to a specific page. Also can select multiple pages to which a tool should be applied, similarly to the way multiple files may be selected in File Manager.

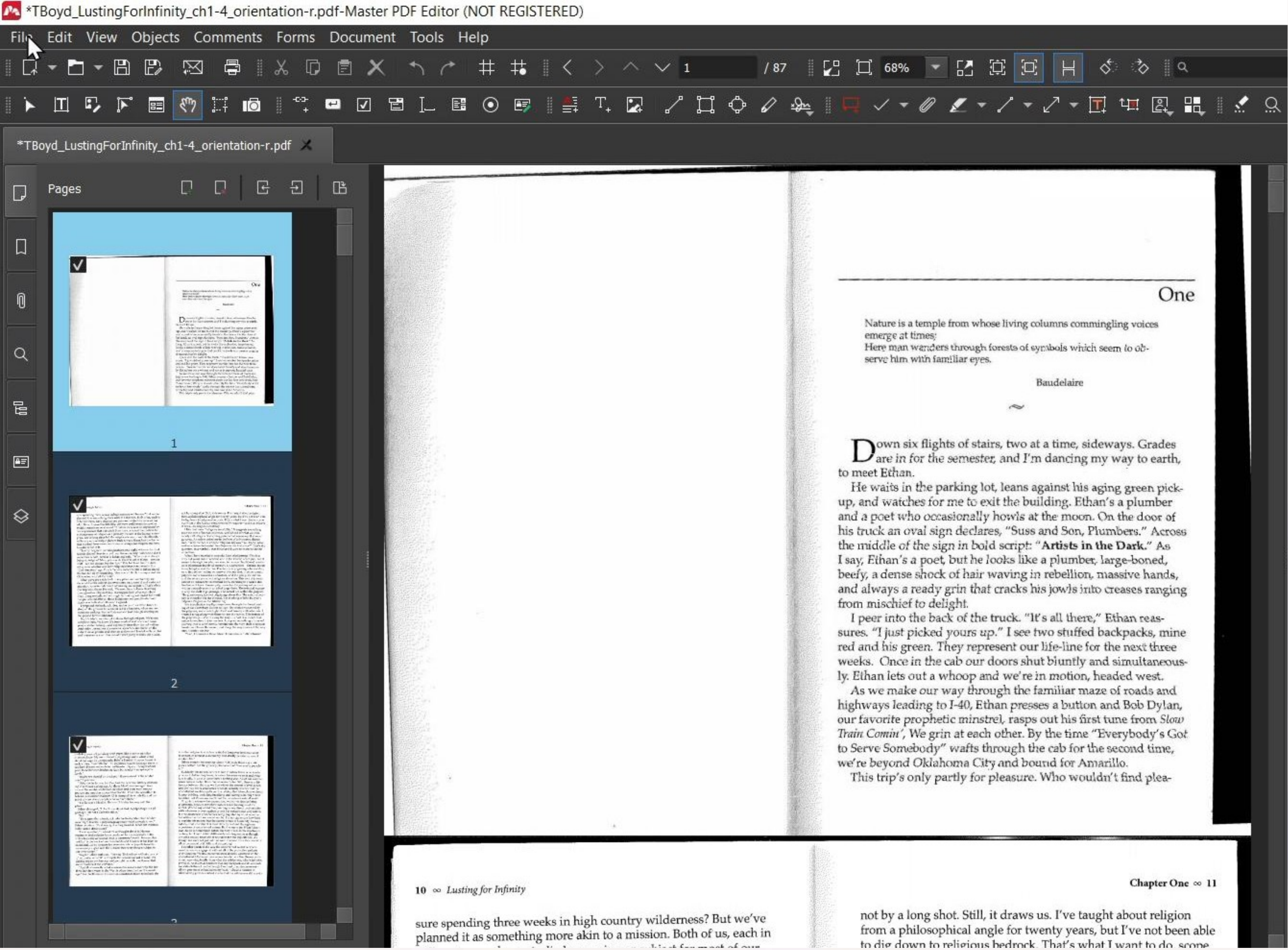
In this case all pages need to be rotated counter-clockwise by 90° to be readable.


TBoyd_LustingForInfinity_ch1-4_orientation-r.pdf-Master PDF Editor
File
Edit
View
Objects
Comments
Forms
Document
Tools
TBoyd_LustingForInfinity_ch1-4_orientation-r.pdf
Pages
Rotate Pages
85

Editing (Rotation)

After selecting all thumbnail images, can click on the "Rotate Pages" icon and than select rotate counter-clockwise by 90°

If you then Save or Save As, the document is saved as rotated. PDF viewers may allow viewing the document rotated, but the actual PDF file still has the wrong orientation.


*TBoyd_LustingForInfinity_ch1-4_orientation-r.pdf-Master PDF Editor (NOT REGISTERED)
File
Edit
View
Objects
Comments
Forms
Document
Tools
Help
1
/ 87
68%
*TBoyd_LustingForInfinity_ch1-4_orientation-r.pdf
Pages
1
2
3

One

Nature is a temple from whose living columns commingling voices
emerge at times;
Here man wanders through forests of symbols which seem to observe him with familiar eyes.

Baudelaire

~

Down six flights of stairs, two at a time, sideways. Grades are in for the semester, and I'm dancing my way to earth, to meet Ethan.

He waits in the parking lot, leans against his aging green pickup, and watches for me to exit the building. Ethan's a plumber and a poet who occasionally howls at the moon. On the door of his truck an oval sign declares, "Suss and Son, Plumbers." Across the middle of the sign in bold script: "**Artists in the Dark.**" As I say, Ethan's a poet, but he looks like a plumber, large-boned, beefy, a dense shock of hair waving in rebellion, massive hands, and always a ready grin that cracks his jowls into creases ranging from mischief to delight.

I peer into the back of the truck. "It's all there," Ethan reassures. "I just picked yours up." I see two stuffed backpacks, mine red and his green. They represent our life-line for the next three weeks. Once in the cab our doors shut bluntly and simultaneously. Ethan lets out a whoop and we're in motion, headed west.

As we make our way through the familiar maze of roads and highways leading to I-40, Ethan presses a button and Bob Dylan, our favorite prophetic minstrel, rasps out his first tune from *Slow Train Comin'*. We grin at each other. By the time "Everybody's Got to Serve Somebody" wafts through the cab for the second time, we're beyond Oklahoma City and bound for Amarillo.

This trip's only partly for pleasure. Who wouldn't find plea-

10 ∞ *Lusting for Infinity*

sure spending three weeks in high country wilderness? But we've planned it as something more akin to a mission. Both of us, each in

Chapter One ∞ 11

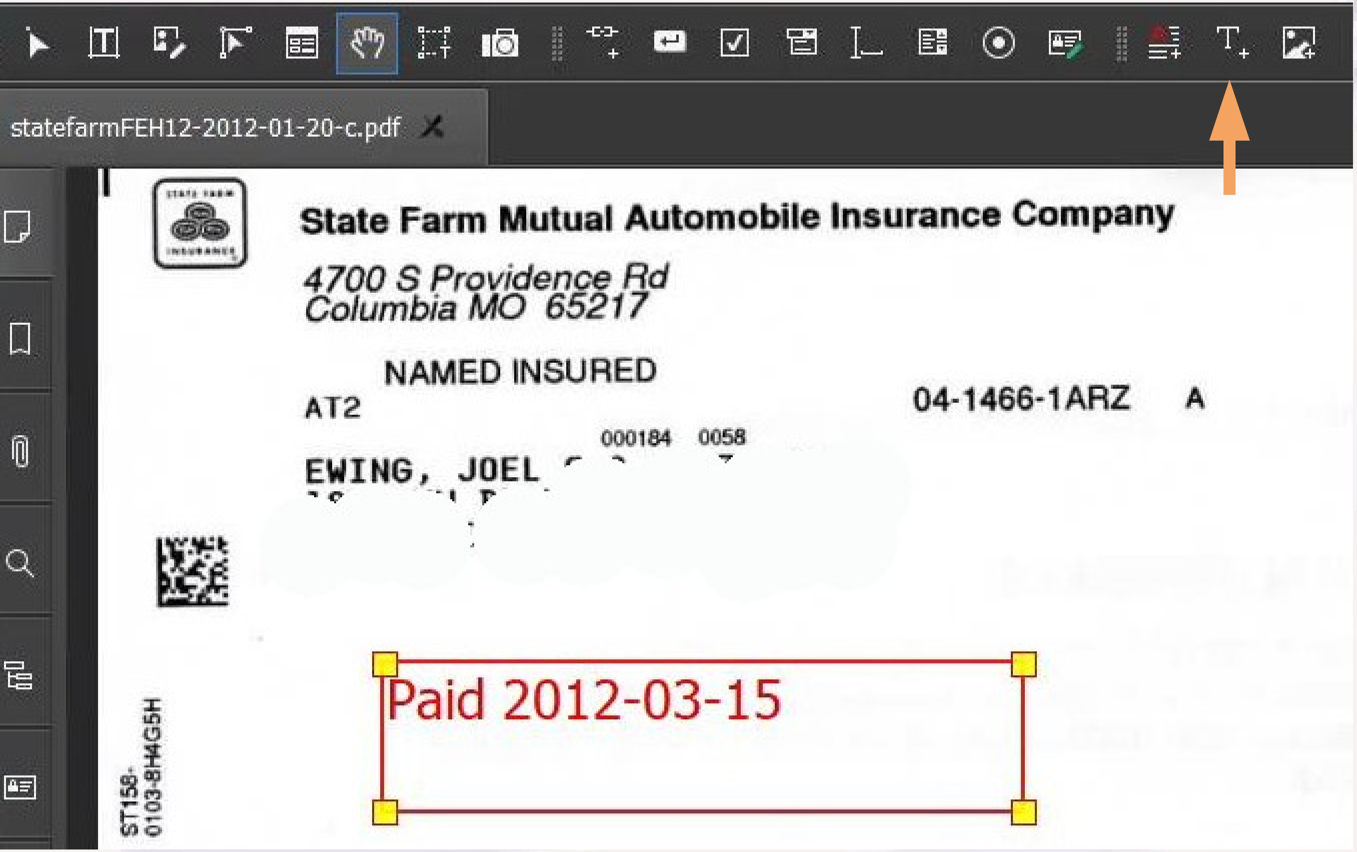
not by a long shot. Still, it draws us. I've taught about religion from a philosophical angle for twenty years, but I've not been able to dig down to religious bedrock. That's what I want to do, scope

After Rotation

All selected pages have been rotated to the correct orientation, including the thumbnail images of the pages.

"Files" "Save As" to save the modified PDF.

Entering Text



statefarmFEH12-2012-01-20-c.pdf

State Farm Mutual Automobile Insurance Company

4700 S Providence Rd
Columbia MO 65217

NAMED INSURED

AT2

04-1466-1ARZ A

000184 0058

EWING, JOEL [redacted]

Paid 2012-03-15

ST158-
0103-8H4G5H

Can enter "Typewriter" (Text) mode, use the cursor to draw a text box, and enter text.

Can choose font, size,color, and optional color and thickness of border (transparent for no border)

Other Useful Functions

- **Can drag & drop thumbnail images to rearrange page order**
- **Can extract selected pages to a new PDF document to split one PDF into multiple PDFs**
- **OCR recognition of text in a PDF with image pages (more on that later)**
- **Other functions as well**

EaseUS PDF Editor

- **- https://pdf.easeus.com**
  - **Free version (functions restricted, product watermark on output)**
  - **Pro $19.47/mo, $49.95/yr, $79.95 lifetime w upgrades per computer**
- **Features**
  - **Convert PDF files to Excel, Word, PowerPoint, images or vice versa**
  - **Edit, OCR, merge, split, compress, create and annotate your PDFs**
  - **Sign, print, encrypt, remove password and add watermark (of your choice) to PDF**

Observations on EaseUS PDF Editor

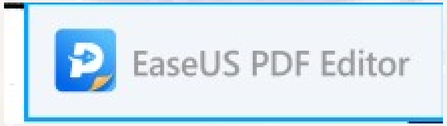
- **Rotate left/right works – can't select arbitrary block of pages, but Even, Odd, and All Pages are included**
- **Split function not supported in Free version (split every n pages) – and other ways to do a split**
- **Can combine multiple files into single file – max of 3 files in Free version**

Observations on EaseUS PDF Editor

- **OCR in Free version limited to 3 pages, and they mean 3 page images: even/odd pages in single PDF page image count as two pages**
  - **Modern fonts, consistent contrast, no distortions work best**
    - **One of the more capable at OCR. Accurately preserved several typos and a spacing error, and even recognized and kept paragraphs in same text block. After OCR, converted to a decent Word document.**
  - **Still can't handle: carbon copies; distorted pages with wavy text; inconsistently formed characters.**

Observations on EaseUS PDF Editor

- **Can move individual pages**
- **Can delete selected pages or a block of pages – which can be used to split a document by repeated editing an original, deleting pages not wanted in each piece, saving, then repeat for each piece.**
- **The "watermark" on each page in Free version very noticeable on each page:**



EaseUS PDF Editor

- **Text insertion works, but not that easy: Can't rotate, not obvious how to delete entire block**
- **Can create a new PDF page from text fields and images**

Observations on EaseUS License

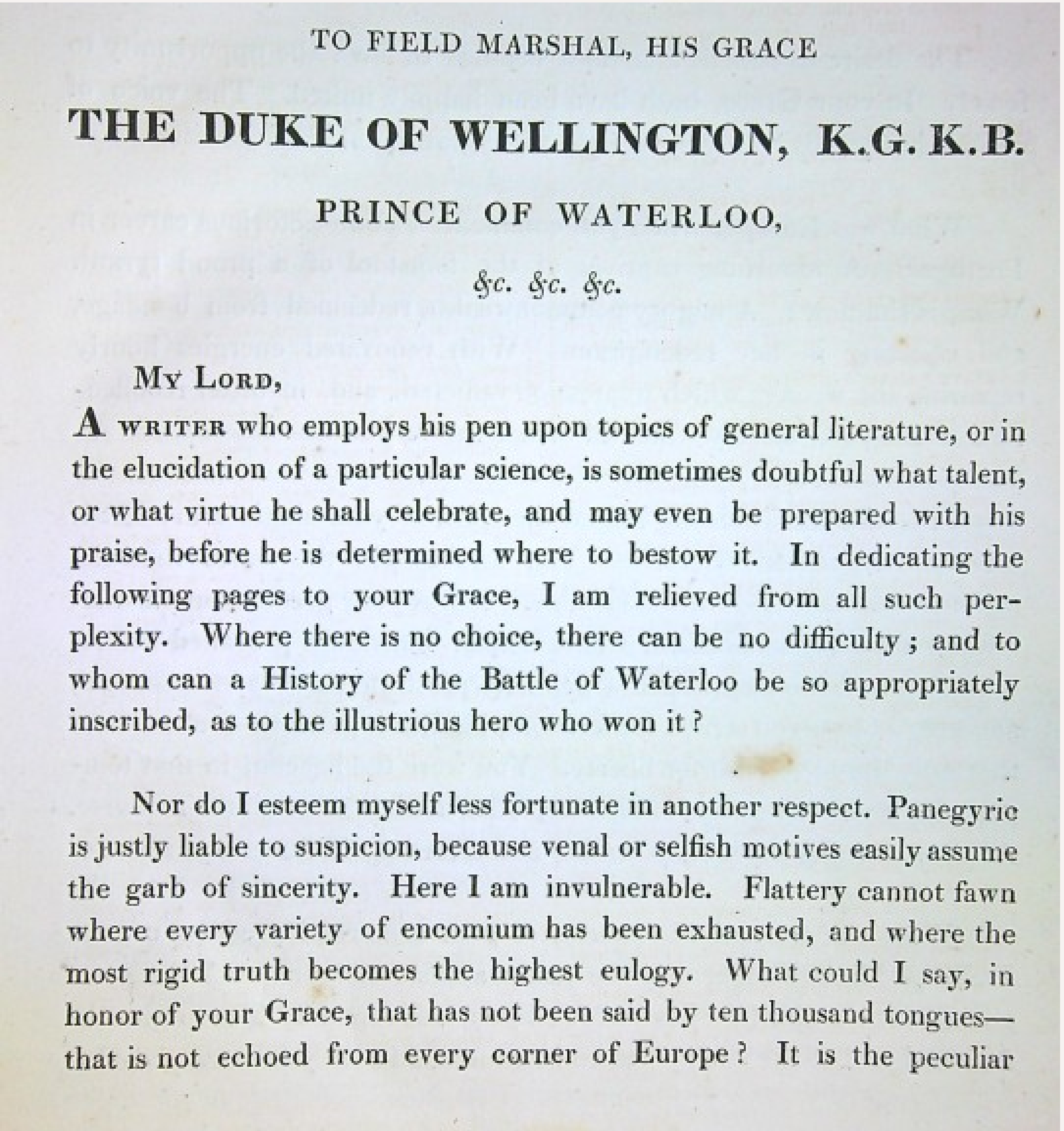
- **EaseUS approach of not letting you have full access until you pay made it more difficult to assess. Code Industries approach of giving full functional access to all features during the trial period proved to be a better strategic approach from the user's standpoint:**
  - **I found more need for the product, so when an unacceptable watermark began at end of the trial period, felt compelled to upgrade to a paid license.**

General Remarks on OCR

- **Results with OCR can be highly variable**
  - **Best results with image PDFs scanned from computer generated documents. Computer-generated fonts tend to be very precise with consistent contrast.**
  - **EaseUS Editor did excellent job on some examples and was also able to produce a very readable MS Word document with consistent font and appearance, but if you add or remove words, still find line wrap issues to work around.**
  - **The Code Industry Editor did a slightly less accurate job of recognizing words, and tended to have problems choosing consistent font styles and sizes. Where lines were distorted and wavy, Code actually did better at recognizing isolated words but jumbled up the word order in sentences. This would be better for text searches, but little use if goal is a good MS Word version.**
  - **In cases where OCR failed badly, a fast typist could spend more time trying to edit out the mistakes than just manually retyping the document.**

Original Page

Excerpt from a very old book, too large & fragile to use a flatbed scanner. PDF created from an overhead-camera scanner:
"An Historical Account of The Campaign in the Netherlands, in 1815…" by William Mudford, published in 1817 – events surrounding the Battle of Waterloo and the defeat of Napoleon by Wellington's forces.



TO FIELD MARSHAL, HIS GRACE

THE DUKE OF WELLINGTON, K.G. K.B.

PRINCE OF WATERLOO,

&c. &c. &c.

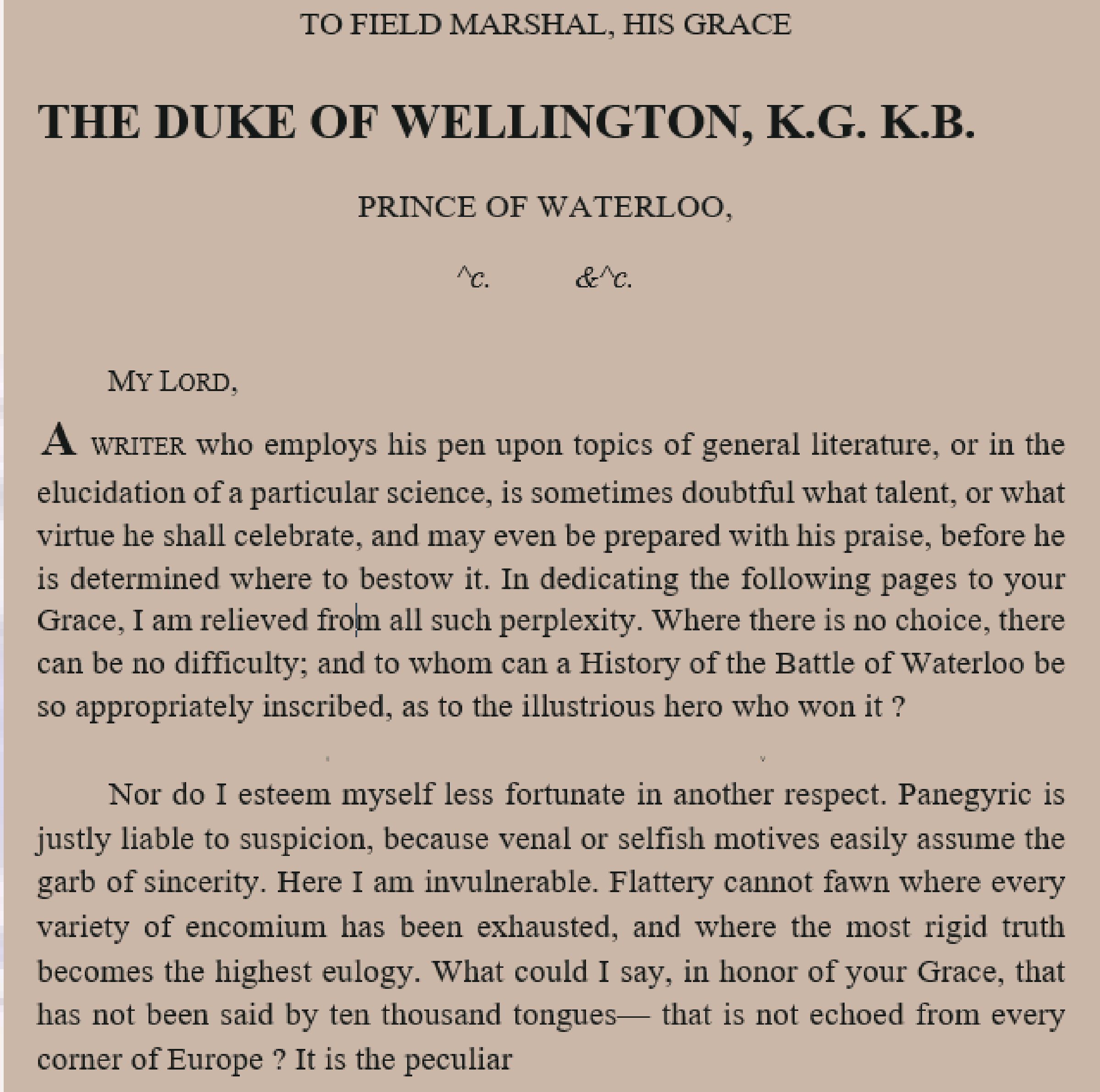
MY LORD,

A WRITER who employs his pen upon topics of general literature, or in the elucidation of a particular science, is sometimes doubtful what talent, or what virtue he shall celebrate, and may even be prepared with his praise, before he is determined where to bestow it. In dedicating the following pages to your Grace, I am relieved from all such perplexity. Where there is no choice, there can be no difficulty; and to whom can a History of the Battle of Waterloo be so appropriately inscribed, as to the illustrious hero who won it?

Nor do I esteem myself less fortunate in another respect. Panegyric is justly liable to suspicion, because venal or selfish motives easily assume the garb of sincerity. Here I am invulnerable. Flattery cannot fawn where every variety of encomium has been exhausted, and where the most rigid truth becomes the highest eulogy. What could I say, in honor of your Grace, that has not been said by ten thousand tongues—that is not echoed from every corner of Europe? It is the peculiar

EaseUS Editor

- **The only obvious OCR failures were for the now-obsolete "&c." notation for "etc."**
- **EaseUS Editor even seemed to recognize the page paper had browned with age.**
- **It preserved the unfamiliar custom of extra space before some punctuation marks (" ?")**



TO FIELD MARSHAL, HIS GRACE

THE DUKE OF WELLINGTON, K.G. K.B.

PRINCE OF WATERLOO,

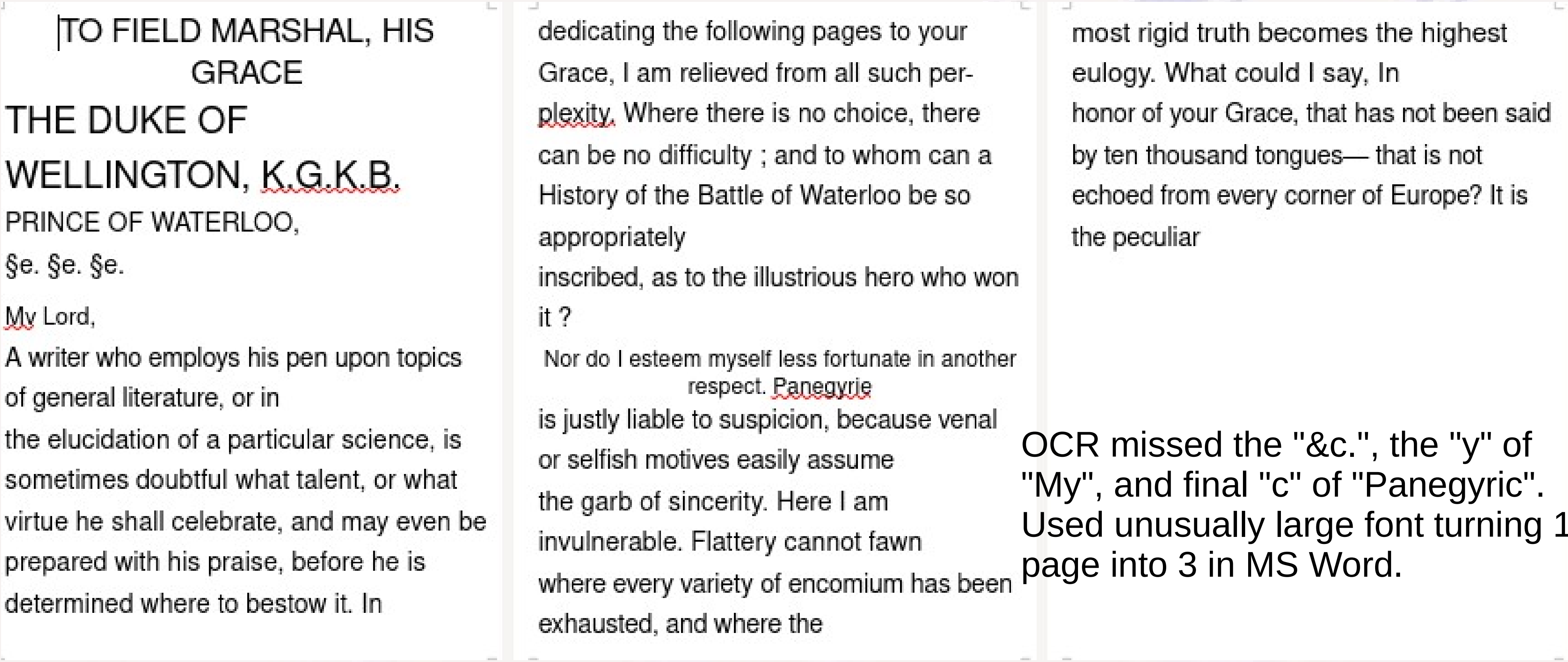
^c. *&^c.*

MY LORD,

A WRITER who employs his pen upon topics of general literature, or in the elucidation of a particular science, is sometimes doubtful what talent, or what virtue he shall celebrate, and may even be prepared with his praise, before he is determined where to bestow it. In dedicating the following pages to your Grace, I am relieved from all such perplexity. Where there is no choice, there can be no difficulty; and to whom can a History of the Battle of Waterloo be so appropriately inscribed, as to the illustrious hero who won it ?

Nor do I esteem myself less fortunate in another respect. Panegyric is justly liable to suspicion, because venal or selfish motives easily assume the garb of sincerity. Here I am invulnerable. Flattery cannot fawn where every variety of encomium has been exhausted, and where the most rigid truth becomes the highest eulogy. What could I say, in honor of your Grace, that has not been said by ten thousand tongues— that is not echoed from every corner of Europe ? It is the peculiar

Core Industry Editor OCR



TO FIELD MARSHAL, HIS GRACE

THE DUKE OF

WELLINGTON, K.G.K.B.

PRINCE OF WATERLOO,

§e. §e. §e.

Mv Lord,

A writer who employs his pen upon topics of general literature, or in

the elucidation of a particular science, is sometimes doubtful what talent, or what virtue he shall celebrate, and may even be prepared with his praise, before he is determined where to bestow it. In

dedicating the following pages to your Grace, I am relieved from all such perplexity. Where there is no choice, there can be no difficulty ; and to whom can a History of the Battle of Waterloo be so appropriately

inscribed, as to the illustrious hero who won it ?

Nor do I esteem myself less fortunate in another respect. Panegyrie

is justly liable to suspicion, because venal or selfish motives easily assume

the garb of sincerity. Here I am invulnerable. Flattery cannot fawn

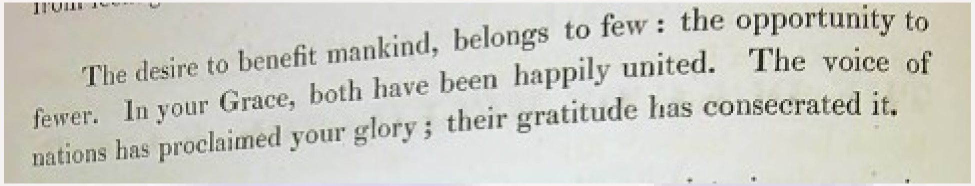
where every variety of encomium has been exhausted, and where the

most rigid truth becomes the highest eulogy. What could I say, In

honor of your Grace, that has not been said by ten thousand tongues— that is not echoed from every corner of Europe? It is the peculiar

OCR missed the "&c.", the "y" of "My", and final "c" of "Panegyric". Used unusually large font turning 1 page into 3 in MS Word.

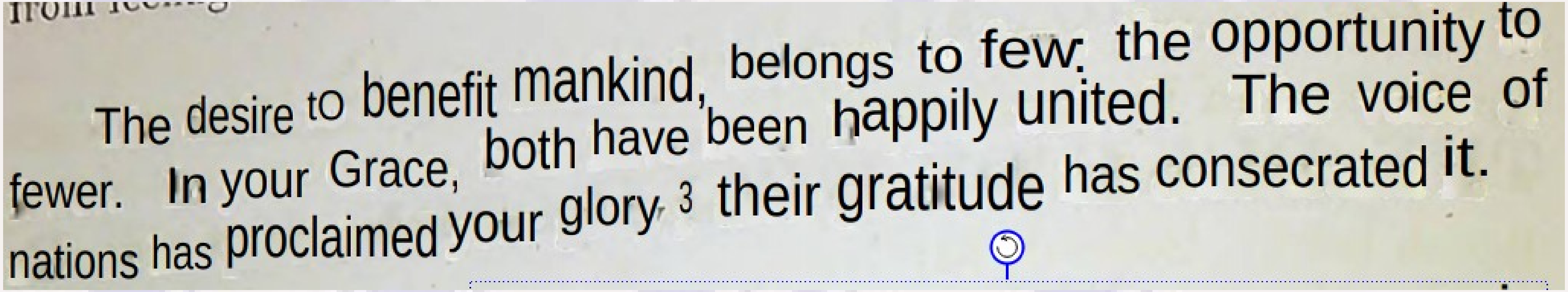
OCR Problem Cases



from feeling

The desire to benefit mankind, belongs to few : the opportunity to fewer. In your Grace, both have been happily united. The voice of nations has proclaimed your glory ; their gratitude has consecrated it.

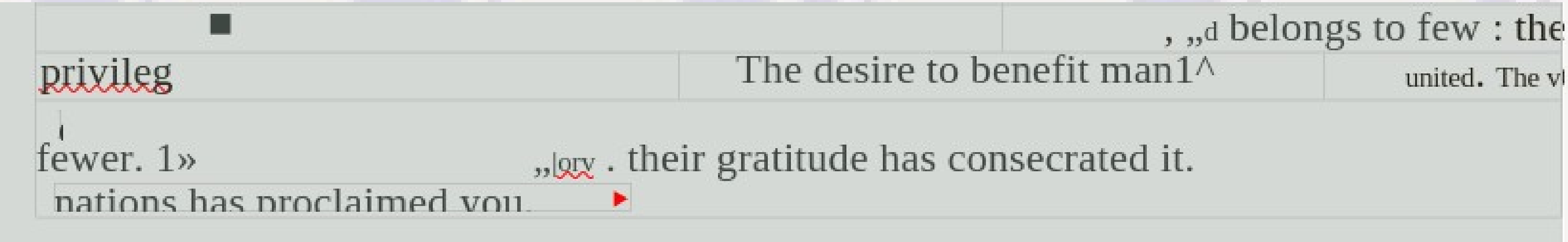
A challenge – scanned page distorted.
Core Industry only missed 1 word
(h appily) -- but when saved as DOCX,
word order very confused



from feeling

The desire to benefit mankind, belongs to few: the opportunity to fewer. In your Grace, both have been happily united. The voice of nations has proclaimed your glory 3 their gratitude has consecrated it.

EaseUS OCR missed 13 words



, ,,d belongs to few : the

privileg

The desire to benefit man1^

united. The v

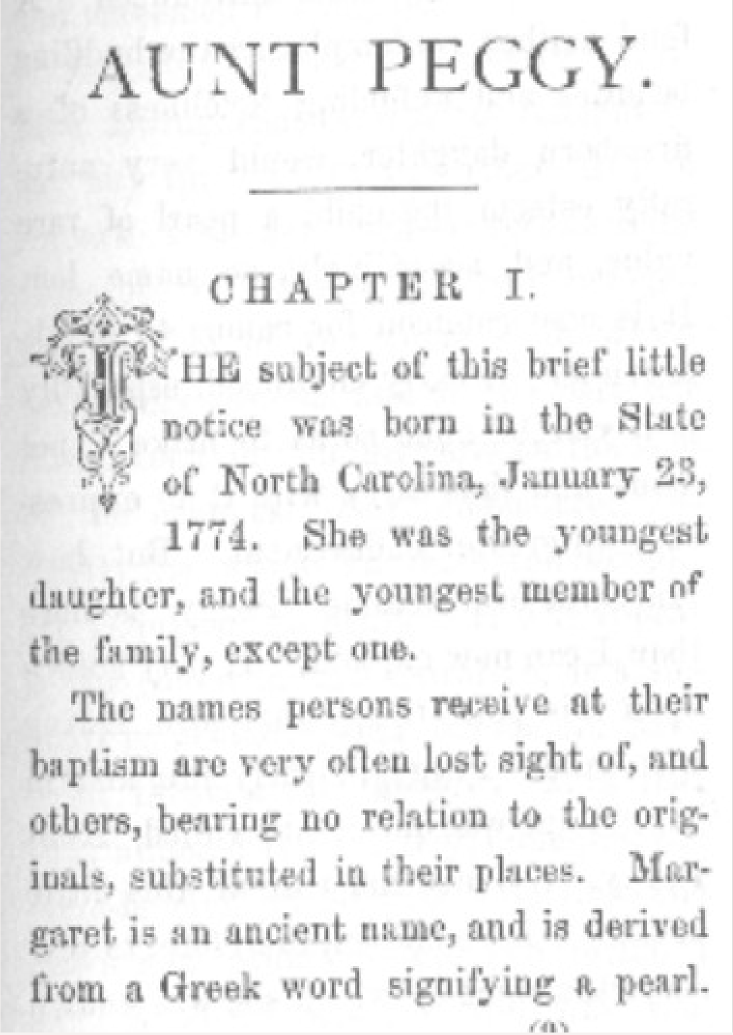
fewer. 1»

,,lorv . their gratitude has consecrated it.

nations has proclaimed you.

OCR Problem Cases

- **Original had both under-inked and over-inked characters, and this was scan of a poor-contrast photocopy reprint of that original.**
- **At best, only 25% of words were correctly recognized**



AUNT PEGGY.

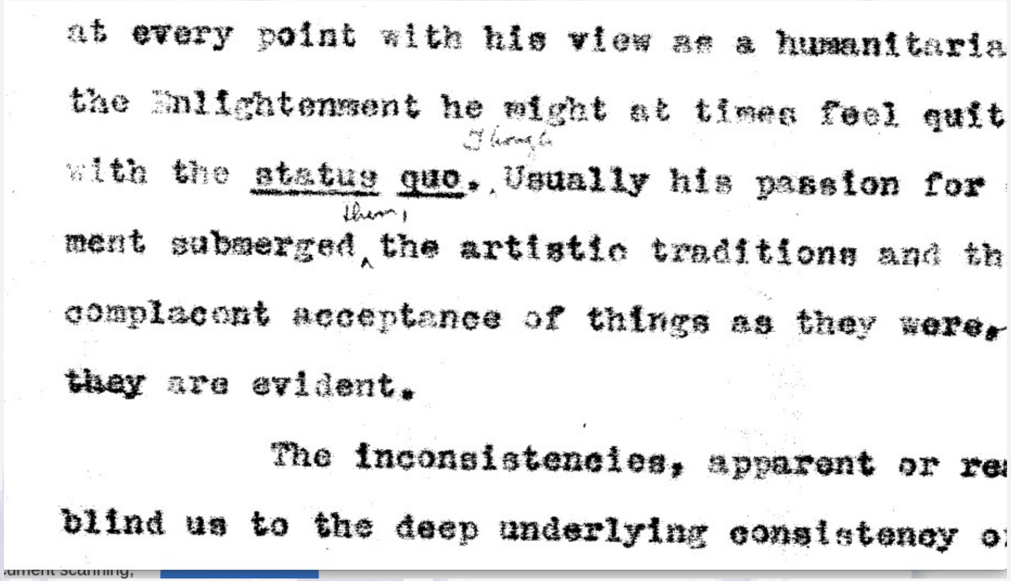
CHAPTER I.

THE subject of this brief little notice was born in the State of North Carolina, January 23, 1774. She was the youngest daughter, and the youngest member of the family, except one.

The names persons receive at their baptism are very often lost sight of, and others, bearing no relation to the originals, substituted in their places. Margaret is an ancient name, and is derived from a Greek word signifying a pearl.

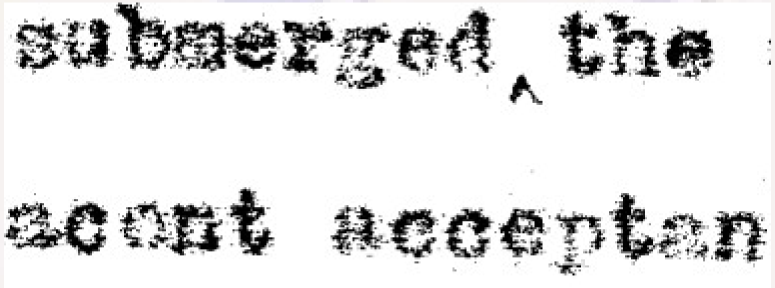
OCR Problem Cases

- **A carbon copy of manually typed 1937 thesis. Characters are mix of smudged, too light, too dark, parts of the letter missing, with hand-written notes.**
- **At best, maybe 1 or 2 words per page were correctly recognized**



at every point with his view as a humanitaria
the Enlightenment he might at times feel quit
Though
with the status quo. Usually his passion for
them,
ment submerged the artistic traditions and th
complacent acceptance of things as they were,
they are evident.

The inconsistencies, apparent or rea
blind us to the deep underlying consistency o



submerged the

acent acceptan

PDF Gear

- **PDF Gear - https://www.pdfgear.com**
  - **Free On-Line tools: Read, edit, convert, split/merge, and sign PDF files**
  - **Currently-free (future plans?) Downloadable utility for Windows, Mac, iOS, Android, ChromeOS. Windows version said to run in Linux using Wine support. https://www.pdfgear.com/pdfgear-for-windows/**
    - **Has a ChatGPT Copilot help component (requires Internet)**
  - **Worth looking at, but unclear how they support the "free" service.**

Windows On-Line-Only PDF Tools

- **If you feel comfortable putting your documents on the 3rd-party website, there are a number of free options available.**
  - **Makes documents more vulnerable.   Might even be used to train AI software.**
  - **Would not be acceptable for documents that have legal privacy requirements that restrict disclosure.**

Windows On-Line-Only PDF Tools

- **Sejda PDF editor - Most versatile - online only**
- **Online Google Docs -**
  - **Can convert from PDF to Google docs format, edit, and save as PDF, bug**
  - **Has some formatting issues**

Windows On-Line-Only PDF Tools

- **PDF Candy -** B**est-overall free On-Line editor**
  - **https://pdfcandy.com & https://pdfcandy.com/download.html**
  - **Unrestricted edit, mark up, and annotate text**
  - **Full-function access to all tools of non-free version, but only 1 usage per hour of those features**
    - **Watermarking documents; adding, rearranging, and splitting pages; extracting images; editing file metadata; cropping and resizing pages; password protecting PDFs; converting to/from variety of formats**
  - **Unrestricted use of "advanced" tools: $6/mo or $48/yr**
  - **Lifetime unrestricted use plus desktop editor: $99**

Linux PDF Tools

- **Linux has free tools for merging, splitting, reordering, rotating PDF pages. There are also command-line tools ("convert") that can do batch conversion of PDF to page images and images to multi-page PDFs.**
- **To get decent user interfaces to deal with text fields inside a PDF page may require a non-free tool. Code Industry Master PDF Editor is also available for Linux.**

Questions?