

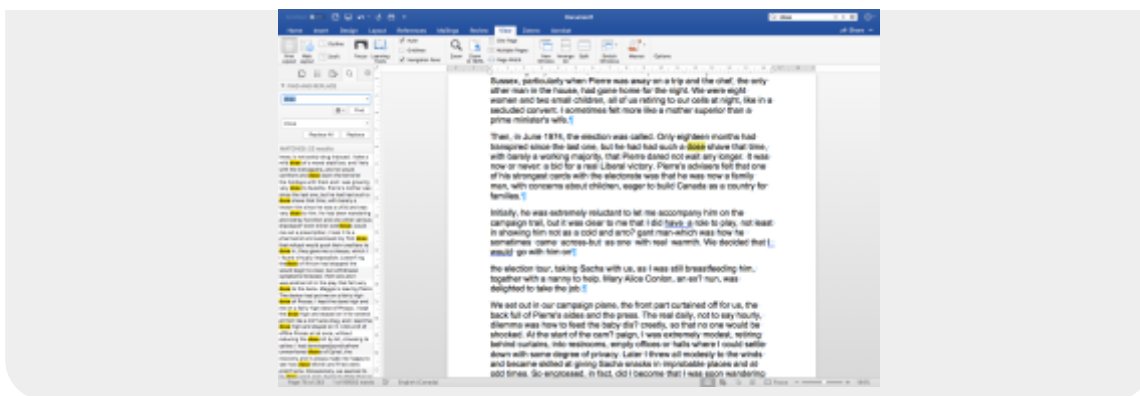
Table of Contents

Find & Replace

1

Q&A

4



Since all of the matches are listed in the sidebar, you will be able to go through each one and see which ones need fixing and which ones can be left alone.

This is not a perfect solution; it is impossible to be aware of every possible error that OCR might make from an older PDF, but it can help streamline some parts of the process. Scanning the entire text will still be necessary (like for the "13"s that become "B"s. Searching for a capital B will generally not be the best way to recognize the error, as there will likely be hundreds of Bs in the text) but using Find & Replace can help you solve some more readily searchable issues.

PROBLEM: There are newlines/line breaks (↵) instead of paragraph marks (¶).

SOLUTION: Find and remove all line breaks and replace with a single paragraph break.

Find: ^m

Replace with: ^p

In LibreOffice, replace all \n with \p to convert them to paragraphs.

PROBLEM: There are extra paragraph breaks. We want to keep the real paragraph breaks and remove the fake extra paragraph breaks.

SOLUTION: Use MS Word's find and replace to remove the extra paragraph breaks using special Word symbols.

Find: ^p^p (you can also search for more than 2 paragraph breaks, i.e. ^p^p^p)

Replace with: ^p

PROBLEM: Running headers. Example, where the first three numbers and the three numbers after the filename is the page number: 231(paragraph break)MacG_9781770494220_5p_all_r1.indd 231(paragraph break)10/27/14 11:56 AM(paragraph break)

SOLUTION: Without using wildcards:

Find: ^#^#^#^pMacG_9781770494220_5p_all_r1.indd
^#^#^#^p10/27/14 11:56 AM^p

Replace with: nothing. If you're doing a paginated title, replace with page breaks.

You will need to remove one of the ^# at the beginning and after the .indd to remove it for 2 digit page numbers, and one last time for single digit page numbers. The following screenshot is an example with a 1-digit page number (see below), followed by the command used to isolate all such instances.



Find:

^#^pMacG_9781770494220_5p_all_r1.indd
^#^p10/27/14 11:56 AM^p

You will also need to do it with the leading ^#^p to catch the footer text that do not have any page numbers with it.

PROBLEM: After conversion, your etext document ended up with a number of section breaks throughout the text.

SOLUTION: Find and remove all section breaks and replace with nothing.

Find: ^b Replace: [Leave blank]

One additional instance in which you might need to remove section breaks can be found on the [Footnotes and Endnotes page](#).

Q&A

Q: I have a book that has a vast number of subheadings for each chapter, and the book itself is very long. Right now all the subheadings are set at H3, but should be H2. Is there a quick way to do this? Note: I can not delete Heading 3 Style. I am currently going through and manually changing them. I am changing the Chapters to H1 and the subheadings to H2.

A: You can use Advanced Find and Replace to search for all text that is styled as Heading 3 and replace that with Heading 2. Here is a screenshot of how that Find & Replace looks on my Mac:

[findreplaceheadings.png](#)

From:
<http://bcl.wiki.libraries.coop/> - BC Libraries Coop wiki

Permanent link:
<http://bcl.wiki.libraries.coop/doku.php?id=public:nnels:etext:find-and-replace&rev=1556201343>

Last update: 2019/04/25 14:09

