Guide for Converting Documents from Word to Latex

Citations

Dealing with citations is generally a pain. So, full disclosure, the following process is by no means painless. However, it probably saves time to manually re-creating citations within Endnote. The below assumes that one is working with Word, and either Zotero or Endnote as the reference management software. Similar techniques may work with Mendeley, but I have not investigated the specifics for that program.

1. What you will need
   1. Reference Management Software
      1. Endnote or Zotero
      2. Jabref
         1. If using Endnote, then will need Endnote import filter for importing Jabref citations into Endnote. Available at <https://github.com/JabRef/EndNode-JabRef-filters/blob/master/EndNote%20Import%20From%20JabRef.enf>
            1. Copy this enf file into the Endnote filters folder, see here for a brief description <https://github.com/JabRef/EndNode-JabRef-filters>
2. Steps
   1. If working with Endnote
      1. Setting up citation keys in word with Jabref
         1. First, open Endnote and the word document containing your references. You should be able to see within Endnote that references that are cited within the word document
         2. Export these references (or your entire library) to txt file with Bibtex format
         3. Import that file into Jabref
         4. Generate citation keys for the new database
         5. Export that Jabref library
            1. to an Endnote txt file (necessary for Word citations)

Ensure that citation keys generated by Jabref show up in the Endnote “Label” field

* + - * 1. To a bibtex file (necessary for end use in Latex)
      1. Import that txt file back into (a new, if you so choose) Endnote library using the special Endnote filter provided at the above link
      2. Now, make a new Endnote Export style. It can be any style; we are only using it to make a custom in-text citation style.
         1. Click Edit > Output Styles > Edit \*Some style\*
         2. Click on Templates (within Citations)
         3. Here, we are going to change the citation format so that citations display as Latex code

Change “Citation” to \autocite{Label}

Or if you prefer another latex command, such as \cite, that’s fine too

Change “Citation – Author (Year)” to \textcite{Label}

“Multiple citation separator” (at the bottom of the window) should be left as a comma

* + - 1. Now, choose this new style as your output style within your word document
         1. In text citations should now display as with the above latex commands, followed by the Jabref citation key in curley brackets
    1. Copying from word to latex
       1. Now, you should simply be able to copy Word text to your latex file
       2. Note, however, that before the latex file will properly display your references, you need to do some preparatory work in the latex file, described in a later section
  1. If working with Zotero
     1. Install the addon Better Bibtex (<https://retorque.re/zotero-better-bibtex/>), which will ultimately enable us to use a custom CSL (reference formatting) style within word to display citations in latex format
        1. But first, a quick note on the awesome other abilities of better bibtex
           1. Better Bibtex add on also allows you to click, drag, and drop (or copy paste) citations from Zotero into pretty much any program as far as I can tell

E.g., into word, TexStudio, Notepad, online Overleaf

* + - * 1. So this makes inserting latex citations into latex documents as you go relatively easy. But what if you need to search for multiple distinct citations? Rather than clicking and dragging a bunch of times, Better bibtex enables you to use your favorite web browser to open a Zotero search bar

For example, with Zotero open, go to the following link:

<http://localhost:23119/better-bibtex/cayw?format=biblatex&clipboard=yes>

A Zotero search bar should open, from which you can access your references. Hitting enter will display latex formatted citations on the screen. With clipboard=yes, this text is automatically copied to your clipboard. Simply refresh the browser window to reopen the search bar for more references. Note that these references must be in your local Overleaf/TexStudio/Latex Editor bib file in order to display in the final pdf

See the link <https://retorque.re/zotero-better-bibtex/citing/cayw/> for further options on customizing the local host link above (e.g., for markdown citations)

* + 1. After installing Better Bibtex, it will generate unique citation keys for your library
       1. You can change the rules for these citation keys in Zotero preferences > Better BibTex > Citation Keys
          1. <https://retorque.re/zotero-better-bibtex/citing/>
    2. After Better Bibtex generates unique citation keys for your library, go to Preferences > Advanced > Config Editor
       1. Accept the risk
       2. Change setting citeprocNoteCitekey to True, which allows the citation keys to be saved within the notes section of the Zotero refs (which is necessary for the in-text citation modification)
    3. Next, create a new output csl style
       1. Preferences > Cite > Style Editor
       2. Copy paste the CSL style provided at the end of this document into the style editor
       3. Save it as a new style in the Zotero styles folder
          1. Name the file with the same entry as the “ID” field of the CSL file, see below
    4. Working from Word, choose this new style as your document reference style
       1. Now, as for endnote, in text citations will be displayed as latex commands and can be copied pasted into latex
       2. However, if you use Zotero’s “Suppress Author” option, or add either prefix or suffix text to in-text citations, then this style will produce incorrectly formatted citations.
          1. Specifically, for the Suppress Author situation, a normal citation will be inserted into latex. So your sentence would read (if you copied straight from Word), “Authors (Authors 1990) argued ….”

Unfortunately this means that you would have to search your document for these sorts of errors

I have written a Macro to fix this problem, but not the next issue

* + - * 1. For the prefix or suffix situation, the latex citation would be formatted incorrectly and produce an error, resulting in no inputted citation. These may be easier to find via log file information or error checking software in your favorite latex GUI.
  1. Getting citations in Latex
     1. Latex first of all requires that all citations to be used in the document are included in a .bib file, which is in the same directory as the current .tex file
        1. To get references from a word document to that bib file
           1. If you used Endnote

Move the bibtex file (that was exported from Jabref) to the folder where you latex file is

* + - * 1. If you used Zotero

Export your library to a bibtex file and move that file to your latex folder

Or, if you only want to export the citations contained within the word document, first

Extract those references from the document [here](https://rintze.zelle.me/ref-extractor/) (this is necessary b/c Zotero does not currently automatically create a group for citations included in a Open or Microsoft word document)

Import those references into Zotero

Create a collection from imported references

Merge duplicates (if desired)

Export the collection to a bibtex file

* + 1. Next, you need to add the following code to your preamble
       1. \usepackage[backend=biber, style = authoryear, sorting = nyt,sortcites]{biblatex}
          1. Sorting and sortcites are not strictly necessary
       2. \addbibresource{name-of-reference-file.bib}
          1. This code tells latex where to look for the bib file. Note that there are no quotes around the file and the file name has no spaces. I found that file names with spaces would not work
    2. Wherever you want the bibliography to be inserted in the document, add
       1. \printbibliography[title = References]
          1. Replace “References” with whatever title you wish

Overall Formatting

1. Can convert Word documents to LaTex documents via Pandoc, and vice versa
   1. Going from Word to LaTex will not preserve everything
      1. But it may give you a head start in terms of some of the formatting
2. Install pandoc
   1. <https://pandoc.org/>
3. To convert from Word to Latex, <https://jabranham.com/blog/2016/11/using-pandoc-export-to-word/>
   1. Open a terminal or powershell window in the folder that includes your .tex file
      1. E.g., on windows, hold shift and right click, which will give you the relevant option
   2. Into the terminal/powershell window (I have only tested this in powershell), use the following command
      1. pandoc --wrap=none my-document.docx -o my-document.tex
4. To convert from Latex to word
   1. There are lots of resources online providing instructions on how to convert latex to word, usually via pandoc (e.g., <https://abnormaldata.wordpress.com/2014/11/17/converting-latex-to-word/>)
   2. In the simplest case, you can simply reverse the code use to generate word from latex
      1. pandoc mydocument.tex -o mydocument.docx
   3. However, this will not preserve formatting well
      1. The short answer here is that pandoc will require
         1. A header tex file
            1. One [website](http://karthur.org/2015/custom-citation-styles-in-latex.html) that I have found notes that Pandoc expects that the latex file to be converted does not contain preamble text (e.g., loading packages)
            2. According to this website, the preamble text should be included in a separate tex file. See the linked website
         2. A latex template file
            1. This latex temple will contain any formatting information that is not in the preamble.

CSL style (Search for “autocite” if you want to see where you should modify code to change the used latex syntax)

Note that I have changed the title field, line 166, and the id field, line 167. Whenever you modify a csl file, you must modify these fields if you want to use both the original and modified csl file

<?xml version="1.0" encoding="utf-8"?>

<style xmlns="http://purl.org/net/xbiblio/csl" class="in-text" version="1.0" demote-non-dropping-particle="sort-only" default-locale="en-US">

<!-- This style was edited with the Visual CSL Editor (http://editor.citationstyles.org/visualEditor/) -->

<info>

<title>Word to Latex</title>

<id>http://www.zotero.org/styles/better-bibtex-citekeys-word-to-latex</id>

<link href="http://www.zotero.org/styles/better-bibtex-citekeys" rel="self"/>

<link href="http://www.zotero.org/support/bbt\_citekeys" rel="documentation"/>

<author>

<name>Emiliano Heyns</name>

</author>

<category citation-format="author-date"/>

<category field="generic-base"/>

<summary>Style to show BBT citekeys</summary>

<updated>2019-02-27T10:14:13+00:00</updated>

<rights license="http://creativecommons.org/licenses/by-sa/3.0/">This work is licensed under a Creative Commons Attribution-ShareAlike 3.0 License</rights>

</info>

<macro name="author-short">

<names variable="author">

<name form="short" and="text" delimiter=", " initialize-with=". " sort-separator=", "/>

<substitute>

<names variable="editor"/>

<names variable="translator"/>

<text macro="title"/>

</substitute>

</names>

</macro>

<macro name="title">

<text value="&quot;"/>

<text variable="title"/>

<text value="&quot;"/>

</macro>

<macro name="year-date">

<choose>

<if variable="issued">

<date variable="issued">

<date-part name="year"/>

</date>

</if>

<else>

<text term="no date" form="short"/>

</else>

</choose>

</macro>

<citation et-al-min="2" et-al-use-first="1" disambiguate-add-year-suffix="true" disambiguate-add-names="true" disambiguate-add-givenname="true">

<sort>

<key macro="author-short"/>

<key macro="title"/>

<key macro="year-date"/>

</sort>

<layout delimiter=", " prefix="\autocite{" suffix="}">

<group>

<text value=""/>

<text variable="note"/>

<text variable="locator" prefix=", p. "/>

</group>

</layout>

</citation>

<bibliography hanging-indent="true" et-al-min="2" et-al-use-first="1">

<sort>

<key macro="author-short"/>

<key macro="year-date"/>

<key variable="title"/>

</sort>

<layout prefix="{" suffix="}">

<group delimiter=", ">

<text macro="author-short"/>

<text macro="title"/>

<text macro="year-date"/>

</group>

</layout>

</bibliography>

</style> <text macro="year-date"/>

</group>

</layout>

</bibliography>

</style>