

Meta-data extraction comparison by Jennifer Lubke, University of Tennessee

As anyone who has ever pursued doctoral-level study can relate, I had amassed numerous piles of annotated course readings and research materials by the second year of my program in Reading Education. Then, I acquired a tablet computer (iPad). Inspired by the tablet's e-reading utilities and robust apps for annotating, sharing, and storing PDFs, I decided to go paperless. I realized if I was to succeed in this conversion, it was time to fully commit to a reference management tool.

That summer, I took a course on digital tools for qualitative research and learned about several state-of-the-art alternatives to the proprietary software (EndNote) supported by my institution. By drawing on social media functionality to provide a collaborative platform for public and private research groups, these new tools could potentially alter the entire academic research experience. I remember being impressed with Mendeley, which claims to be the "world's largest crowd-sourced library," and I admired the developer community behind open-source Zotero.

But, admittedly, my choice would ultimately be determined by self-interest. I wanted a reference manager that would integrate seamlessly with my search and retrieval of online resources and that would support my new paperless reading and researching routines. Which of the tools would be most effective for reigning in and organizing my growing personal library of PDFs?

I was not alone. Media specialist and blogger Aaron Tay (2010) also wondered about the proliferation of Web 2.0 reference managers, asking: "How good are they at figuring out citations from PDFs?" Tay ran a series of "non-scientific tests" to see how well EndNote, Mendeley, WizFolio, and Zotero ingested a collection of 10 PDFs. In the end, none of the programs worked perfectly, and it seemed Mendeley and Zotero delivered a neck-and-neck performance.

Intrigued, I tried my own mini-metadata experiment using articles representing a range of years (from 1989 to 2011) and a range of PDF versions, with and without DOIs. I focused on Mendeley and Zotero, using both applications' capability for inserting formatted citations by drag-and-drop technology. Zotero and Mendeley provide multiple avenues for adding content, from "single-click" technology to manual entry. I put the single-click option to the test because, in my mind, this functionality would make or break the tool.

Even with a number of documented updates and improvements between the time Tay conducted his metadata extraction and the time I conducted mine, my experiment also yielded inconclusive results. For figuring out citations from imported PDFs, both Mendeley and Zotero are a little buggy. Here are a few examples:

Manually formatted APA citation	Mendeley citation	Zotero citation
Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. <i>Educational Researcher</i> , 18(1), 32-42.	Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. <i>Educational Researcher</i> , 18(1), 32(missing page range). doi:10.2307/1176008	Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. <i>Educational researcher</i> , 18(1), 32(missing page range).
Mishra, P. & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. <i>Teachers College</i>	MISHRA, P., & KOEHLER, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. <i>Teachers</i>	Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. <i>Teachers college record</i> ,

<i>Record</i> , 108(6), 1017-1054.	<i>College Record</i> , 108(6), 1017-1054. doi:10.1111/j.1467-9620.2006.00684.x	108(6), 1017(missing page range).
------------------------------------	--	-----------------------------------

The obvious implication is that researchers will have to consider factors other than pure bibliographic metadata extraction when selecting a citation management system. Do you prefer the way Mendeley automatically locates and adds the DOIs to citations? Or, do you especially like Zotero's tendency to double-space citations? The drag-and-drop technology performs flawlessly within both applications, but you may prefer (as I do) Zotero's split-screen format that integrates with your browser.

Reference:

Tay, A. (2010). Extracting metadata from PDFs: Comparing EndNote, Mendeley, Zotero & WizFolio. *Musings about Librarianship blog*. Retrieved August 22, 2012, from <http://musingsaboutlibrarianship.blogspot.com/2010/07/extracting-metadata-from-pdfs-comparing.html>