ReTracker: Actively and Automatically Matching Retraction Metadata in Zotero

Yi-Yun Cheng, Nikolaus Parulian, Tzu-Kun Hsiao, Ly Dinh, Janina Sarol, Jodi Schneider
School of Information Sciences, University of Illinois at Urbana Champaign, USA
{yiyunyc2, nnp2, tkhsiao2, dinh4, mjsarol, jodi}@illinois.edu

ABSTRACT

Retraction removes seriously flawed papers from the scientific literature. However, even papers retracted for scientific fraud continue to be cited and used as valid after their retraction. Retracted papers are inadequately identified on publisher pages and in scholarly databases, and scholars' personal libraries frequently contain retracted papers. To address this, we are developing a tool called *ReTracker* (https://github.com/nikolausn/ReTrackers) that automatically checks a user's Zotero library for retracted articles, and adds retraction status as a new metadata field directly in the library. In this paper, we present the current version of *ReTracker*, which automatically flags retracted articles from PubMed. We describe how we have iteratively improved *ReTracker*'s matching performance through its initial two versions. Our tests show that the current version of *ReTracker* is able to flag retracted articles from PubMed with high precision and recall, and to distinguish retracted articles from articles about retraction. In its current state, *ReTracker* can actively and automatically bring retraction metadata into Zotero, and in future work we will test its usability with scholars.

KEYWORDS

Retraction; Scholarly Communication; Publishing Ethics

ASIS&T THESAURUS

Information Resources Management; Scholarly Publishing

INTRODUCTION

Retraction removes seriously flawed papers from the scientific literature. Reasons for retraction include honest error, scientific fraud, and plagiarism [Bar-Ilan and Halevi, 2017; Da Silva and Bornemann-Cimenti, 2017]. The Committee on Publication Ethics, an industry group that develops best practices for scholarly publishing, says that "Prompt retraction should minimize the number of researchers who cite the erroneous work, act on its findings or draw incorrect conclusions" [COPE, 2009].

However, papers continue to be cited after retraction, sometimes at a higher rate than before retraction [Bar-Ilan and Halevi, 2017; Da Silva and Bornemann-Cimenti, 2017; Dong et al., 2005]. Even fraudulent papers continue to accrue citations. For instance, in the five years following their retraction for fabricating data, 20 papers on fraudulent pain research by Scott Reuben were cited a total of 267 times in 122 different journals, yet only a quarter mentioned the retraction [Bornemann-Cimenti et al., 2016].

Other studies have found that post-retraction citation is mostly positive, and rarely signals the retraction. For instance, bibliographies almost never indicated the retractions in Bar-Ilan & Halevi's study of 238 post-retraction citations to 15 papers across disciplines, and only 12 (5%) of these 238 post-retraction citations were negative [Bar-Ilan and Halevi, 2017].

In a print environment, medical librarians could disseminate information about retracted papers by flagging print copies and sharing lists of retracted papers [Snodgrass and Pfeifer, 1992], and many did so [Hughes, 1998]. Retracted papers are inadequately identified on publisher pages and in scholarly databases, and scholars' personal libraries frequently contain retracted papers [Da Silva and Bornemann-Cimenti, 2017; Davis, 2012]. In a 2012 study of 1,779 retracted articles in medicine, only 5% of PDFs indicated the retraction; this raised to only 26% when including landing pages. Of these articles, 75% were available in personal Mendeley libraries [Davis, 2012].

Systems to surveil personal PDF libraries (as well as bibliographies in pre-press manuscripts) have been suggested numerous times (e.g. [Bar-Ilan and Halevi, 2017; Davis, 2012]). A common recommendation from researchers in diverse fields is to flag retracted articles in bibliography management tools [Cosentino and Veríssimo, 2016; Granter and Papke, 2018].

To meet this need, this paper presents a tool called *ReTracker* that automatically detects retracted papers within the popular bibliography management tool, Zotero. *ReTracker*'s development began in 2018; its architecture [Dinh et al., 2019] and open source code (https://github.com/nikolausn/ReTrackers) are publicly available. The goal of this paper is to demonstrate the reliability of *ReTracker*'s approach to matching retracted publications from PubMed, and to describe the development path for the project.

CURRENT TOOLS FOR IDENTIFYING RETRACTION

Retraction Watch Database

The Retraction Watch database[1] initially released in October 2018 [Brainard, 2018], contains the largest repository of retracted papers, reported at 19,303 articles from 1756 through mid-June 2019. The database allows any user to search various access points of retracted papers (including Author, Title, Reasons, etc), limited to 600 results. In our examination of the data, we have found articles that we could not confirm were retracted (e.g., withdrawn pre-press articles without issuing issue retraction notices). Bulk data is available for bibliometric analysis, but the licensing prohibits redistribution of the data.

re-cite

The re-cite website[1] highlights retracted papers in their database, given a list of APA or AMA references. Their database of 3,107 retracted articles published before 2016 was built by extracting articles from Web of Science with "retraction of" in the title, removing articles not categorized as corrections, and filtering out possible false positives. The database's inability to highlight the Wakefield paper [Wakefield et al.,1998], retracted in 2010, suggests incomplete data even for the years covered. The database is not being consistently updated.

CrossMark

A CrossMark button[2] on an HTML, PDF, or EPUB paper enables the reader to "check for updates", when publishers have opted in to this paid CrossRef service. Clicking on the CrossMark button takes the reader to a landing page indicating any updates to the paper, including retractions. One limitation is that few publishers use CrossMark because they must pay fees for each record that is updated or corrected in CrossRef databases.

Open Retractions

Open Retractions[3] enables users to check whether a paper has been retracted, using the paper's DOI as a search query.

However, the tool requires a DOI and is highly dependent on finding the DOI and its retraction status in PubMed and CrossRef. The tool's inability to flag well-known retracted papers such as the Wakefield paper [Wakefield et al., 1998] warrants further improvements.

Limitations of Current Tools

With most of these current tools, readers must check one paper at a time, by searching for its title, author, DOI, or reference, or by clicking on a "check for updates" button in the paper itself. The only tool that checks more than one paper at a time (re-cite) is not updating their records. By contrast, our approach actively and automatically adds the most current retraction metadata from PubMed into Zotero.

RETRACKER TOOL

We have developed a novel tool *ReTracker* that automatically detects retracted papers in the popular bibliography management tool Zotero. Whenever the user downloads or syncs items in their Zotero library, *ReTracker* sends queries to PubMed to identify retracted articles, and flags the retracted ones. Retracted articles are flagged with retraction metadata in red, and users can click on the "[source]" button to navigate to the article's PubMed page (see Figure 1). This automatic design saves users' efforts in repetitively searching for and manually flagging retracted papers from external sources. The tool's infrastructure and features are described in [Dinh et al., 2019]. The current release of the tool and instructions on how to install ReTracker are freely available via our GitHub[4].

We use PubMed as the data fetching point because: (1) It is freely, publicly available, with no restrictions on its redistribution; (2) It has been recognized as a reliable source for finding retracted articles, regardless of minor inconsistent retraction metadata [Bar-Ilan and Halevi, 2017; Schmidt, 2018]. Furthermore, unlike the tools mentioned above, ReTracker does not keep a local database of retracted articles. Our approach of directly querying data from PubMed gets the most up-to-date information on retractions.
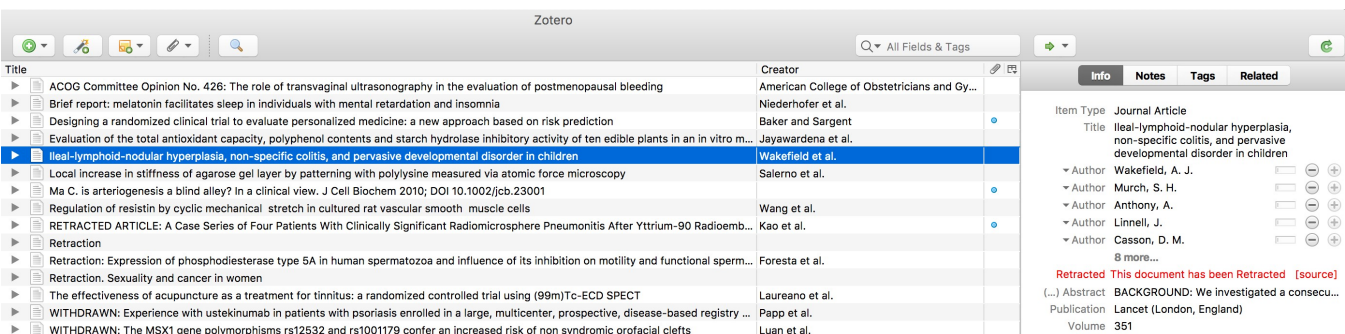

Figure 1. Interface of Zotero with *ReTracker* plugin showing a retracted article with retraction (in red) added by *ReTracker*.

---

[1] Retraction Watch database: http://retractiondatabase.org
[1] re-cite: http://re-cite.org and https://github.com/recite/
[2] CrossMark: https://www.crossref.org/services/crossmark/

[3] Open Retractions: http://openretractions.com/ and https://github.com/fathomlabs/open-retractions/
[4] ReTracker Github repository: https://github.com/nikolausn/ReTrackers

ITERATIVE DEVELOPMENT AND TESTING

ReTracker Version One

The first version, *ReTracker v1*, queries the title of each article in users' Zotero library using PubMed's search box. It iterates the search until an exact match was found. To test the performance of *ReTracker*, we collect test sets in different stages. Our test sets were collected from March 29 to 30, 2019.

Test set one: Retracted articles (R). To evaluate how well *ReTracker v1* flags retracted articles, we tested against Test Set One. Test Set One consists of a representative random sample of 1,250 retracted articles, from the 13,463 articles retrieved from PubMed with publication types "retracted publication" and "retraction of publication" with the following search:

```
"retracted        publication"[Publication    Type]
OR "retraction of publication" [Publication Type]
```

Test set one: Not-retracted articles (NonR + TopicR). Similarly, not-retracted articles are also used to test *v1* to check whether *ReTracker* correctly *avoids* flagging these articles. These were identified by two means:

1. The not-retracted articles (NonR)

27,624,608 results are returned by the following query:

```
("journal        article"[Publication        Type])
NOT ("retracted   publication"[Publication   Type]
OR "retraction of publication"[Publication Type])}
```

2. Articles About Retractions (TopicR)

Articles studying retractions often have the keyword "retraction" in the title and will be hard to distinguish from retracted articles. To identify articles that analyze retractions, we search the MeSH term "*Retraction of Publication as Topic*" *[MAJR]* in PubMed, and find 370 records on such articles.

We sampled 1,000 not-retracted articles from **NonR** and added the 370 articles in **TopicR** to form the test set of not-retracted articles (n=1,370).

V1: Testing result. *ReTracker v1* identified 1,129 of retracted articles (**R**) as true positives, achieving a relatively high recall (0.903) in identifying retracted papers. *ReTracker v1*'s precision, at 0.806, was fair.

For the not-retracted articles, of the 1,000 **NonR** papers sampled, none were flagged as retracted. However, the precision fell short when *ReTracker* encountered articles *studying* retractions (**TopicR**). Out of the 370 papers that PubMed classified as **TopicR**, *ReTracker v1* falsely flagged 271 of them as "retracted" articles because the title or the abstract of the paper contained variations of the word 'retract'. Full results are shown in Table 1. Metrics for calculating accuracy, precision, and recall can be found in the Appendix.

Version	Test Sets	Accuracy	Precision	Recall
ReTracker v1	R, NonR+TopicR	0.850	0.806	0.903
ReTracker v2	R, NonR+TopicR	0.789	0.994	0.561
ReTracker v2	Precise R, NonR+TopicR	0.963	0.999	0.926

Table 1. The performance of the two versions of *ReTracker*

Error Analysis. In our error analysis, we noticed two problems. First, for search queries that return multiple entries, *ReTracker v1* could not always find a single paper. Second, retraction notices were flagged along with retracted papers, because their titles are often similar, e.g. "Retraction Note:" followed by the name of the paper. This motivated us to improve the matching process.

ReTracker Version Two

To mitigate two matching issues—handling multiple search results, and falsely flagging papers studying retractions—we developed a second version of *ReTracker* with an improved matching process. *ReTracker v2 iterates* through results pages with multiple entries until it finds an exact match. *ReTracker v2* also stopped flagging retraction notices of articles.

Test set one. To test the overall performance of *ReTracker v2* we first reused Testing Set One for both retracted and not-retracted articles.

V2: Testing result 1. With the same testing sets as v1, high precision is reported at this stage (0.994), with only 4 of the not-retracted articles flagged as retracted (**NonR+TopicR**). After manual inspection, we found that 3 of these 4 articles (PMID 1471789, 25385480, 15495094) are in fact retracted articles misclassified into the MeSH term "Retraction of Publication as Topic". The remaining article, PMID 7854817, is titled "Fraudulent publication", but is a guest editorial from 1994 that sought to define terms relating to scientific misconduct. *ReTracker v2* still flags it as retracted based on the title of the paper. However, results for this phase show that the accuracy (0.789) and recall (0.561) have significantly dropped from v1. Only 701 out of the 1,250 retracted articles were flagged by *ReTracker*.

Error Analysis. In our error analysis, we discovered that most articles that are *not flagged* were in fact *retraction notices*, suggesting that *ReTracker v2* can flag the original retracted article but avoid flagging the retraction notices. This led us to retest *ReTracker v2* against a more precise pool of retracted articles, (**PreciseR**).

Test set two: Retracted Articles (Precise R). We develop a more optimal search strategy for precision retrieval of retracted articles in PubMed, with the following search:

```
(("retracted      publication"[Publication      Type]
OR"retraction of publication"[Publication Type])
NOT ("Removal     Notice"[TI]    OR    "TEMPORARY
REMOVAL"[TI] OR statement*[Ti] OR editorial*[TI]
OR     retract*[TI]     OR     erratum*[TI]     OR
withdraw*[TI]))
```

With the above search, **PreciseR** consists of a representative random sample of 1,250 articles, from the 7,399 articles retrieved from PubMed.

For not-retracted articles, we reused Test Set One with **NonR + TopicR**.

V2: Testing result 2. Testing *ReTracker v2* against Test Set Two helps us reflect the *true* positives of retracted articles (**PreciseR**). The results have improved substantially, compared to *ReTracker v1* and the testing result 1 of *ReTracker v2*. Out of the 1,250 retracted papers sampled, *ReTracker v2* was able to flag 1,158 of them as retracted. Again, *ReTracker* avoided flagging any of the **NonR**. When we update **TopicR** to only 1 article being flagged (PMID 7854817), the results reflect that *ReTracker v2* has higher accuracy (0.963), precision (0.999), and recall (0.926).

DISCUSSION AND CONCLUSION

The current version of *ReTracker* is able to flag retracted articles with high accuracy, precision, and recall, while not flagging articles that discuss retractions and not flagging retraction notices. *ReTracker*'s current matching process is based on an article's title, given that every article must have a title (whereas not all articles have metadata such as DOIs).

Some of our design choices for *ReTracker* are constrained by the tool development environment and the data licensing. Zotero was chosen not only because it is a popular bibliography management tool, but also since it allows plug-ins such as *ReTracker* to be integrated. For similar reasons, we chose PubMed as *ReTracker*'s external source because it is freely and publicly available for redistribution.

In future development, we would like to (1) expand the *ReTracker* plugin to other bibliography management tool environments (e.g. EndNote, Mendeley, etc.) and (2) incorporate additional external sources for *ReTracker*'s retraction metadata. We also plan to evaluate *ReTracker*'s usability through a user study with Zotero users.

In its current state, *ReTracker* can actively and automatically bring retraction metadata into Zotero. In future releases, when other external sources are aggregated, more rounds of iterative testing and development will be inevitable.

REFERENCES

Bar-Ilan, J. and Halevi, G. (2017). Post retraction citations in context: A case study. *Scientometrics, 113*(1), 547–565.

Brainard, J. (2018). What a massive database of retracted papers reveals about science publishing's 'death penalty'. http://doi.org/10.1126/science.aav8384

Bornemann-Cimenti, H., Szilagyi, I. S., and Sandner-Kiesling, A. (2016). Perpetuation of retracted publications using the example of the Scott S. Reuben case: Incidences, reasons and possible improvements. *Science and Engineering Ethics, 22*(4),1063–1072.

COPE (2009). Guidelines for retracting articles. https://doi.org/10.24318/cope.2019.1.4

Cosentino, A. M. and Veríssimo, D. (2016). Ending the citation of retracted papers. *Conservation Biology, 30*(3), 676–8.

Da Silva, J. A. T. and Bornemann-Cimenti, H. (2017). Why do some retracted papers continue to be cited? *Scientometrics, 110*(1), 365–370.

Davis, P. M. (2012). The persistence of error: a study of retracted articles on the Internet and in personal libraries. *Journal of the Medical Library Association: JMLA, 100*(3), 184–189.

Dinh, L., Cheng, Y.-Y., and Parulian, N. (2019). ReTracker: an open-source plugin for automated and standardized tracking of retracted scholarly publications. In *Proceedings of the 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL 2019)*, 406–407.

Dong, P., Loh, M., and Mondry, A. (2005). The "impact factor" revisited. *Biomedical Digital Libraries, 2*(1), 7.

Granter, S. R. and Papke, D. J. (2018). Opinion: Medical misinformation in the era of Google: Computational approaches to a pervasive problem. *Proceedings of the National Academy of Sciences, 115*(25), 6318–6321.

Hughes, C. (1998). Academic medical libraries' policies and procedures for notifying library users of retracted scientific publications. *Medical Reference Services Quarterly, 17*(2), 37–42.

Schmidt, M. (2018). An analysis of the validity of retraction annotation in PubMed and the Web of Science. *Journal of the Association for Information Science and Technology, 69*(2), 318–328.

Snodgrass, G. L. and Pfeifer, M. P. (1992). The characteristics of medical retraction notices. *Bulletin of the Medical Library Association, 80*(4), 328–334.

Wakefield, A. J., et al. (1998). [RETRACTED:] Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet, 351*(9103), 637–641.

APPENDIX: METRICS*

$$Accuracy = \frac{R_f + NonR_{nf} + TopicR_{nf}}{TotalNumberOfEvaluatedSubjects} \quad (1)$$

where R_f is the number of retracted articles flagged as retracted[2]; $NonR_{nf}$ is the number of not-retracted articles *not* flagged as retracted; $TopicR_{nf}$ is the number of articles about retractions *not* flagged as retracted.

$$Precision = \frac{R_f}{R_f + NonR_f + TopicR_f} \quad (2)$$

where $NonR_f$ is the number of not-retracted articles flagged as retracted; $TopicR_f$ is the number of articles about retractions flagged as retracted.

$$Recall = \frac{R_f}{R_f + R_{nf}} \quad (3)$$

where R_{nf} is the number of retracted articles not flagged as retracted.

* These metrics are written for Test Set One. To adapt to Test Set Two, which uses a different sample pool of retracted articles, all the R_f change to **PreciseR_f**, and all the R_{nf} change to **PreciseR_{nf}**.