Randi Proeschildt and Jodi Schneider
University of Illinois at Urbana-Champaign

Retracted Papers with Incorrect Document Type Indexing in PubMed, Scopus, and Web of Science

**Introduction**

- Finding all retracted publications in a database is important for bibliometric researchers; meanwhile domain scientists may need to eliminate all retracted publications from search results, or to determine whether a particular article is retracted.
- By searching PubMed, Scopus, and Web of Science (WoS) and analyzing the titles of the search results, we evaluate the accuracy of document type classification related to retraction.
- Our goal is to raise awareness of the continuing challenge of quality control issues relating to retracted publications.

**Related Work**

- Donner (2017)
  - Accuracy of document types in WoS
  - Does not study “retraction” nor “retracted publication” document types, which were not introduced to WoS until 2016 (Clarivate Analytics, 2018)
- Schmidt (2018)
  - Quality of retraction indexing—both document types and interlinking of retracted publications and retraction notices

**Method**

To identify possible retracted articles not indexed as such, we:
- Search for “retracted article,” a phrase commonly found in retracted article titles, and refined our search by eliminating retraction-related document types (Table 1).
- Sorted the search results by title, read each title, and identified which were likely to be retracted articles.
- Investigated search results from Scopus, the largest set, by analyzing publishers and DOI-matching against Retraction Watch data (2020) (Figure 2).

**Key findings**

- Most articles were easily identifiable as retracted publications or retraction notices due to common patterns in their titles (Figure 1).
- 8654 items in Scopus lacked retraction-related document types.
- Of these, 8386 were published by the Institute of Electrical and Electronics Engineers (IEEE).
- The IEEE items we were able to obtain retraction dates for ranged from 2007-2017.

**Data Availability**

Data (except the Retraction Watch-matched data) is publicly available at https://osf.io/epd2z/.

**Results**

| Document types | Search 1: “Retracted article” in title, but not indexed with retracted article document type | Num

| Likely retracted articles (percentage) |
|----------------------------------------|---------------------------------|------|
| PubMed (New PubMed) | Retraction of Publication | 61 | 58 (95%) |
| Scopus | Retracted | 8683 | 8654 (99%) |
| Web of Science (All databases, all years) | Retraction Correction, Addition | 80 | 56 (70%) |

**Discussion**

- The large number of search results in Scopus can be tied to a large scale retraction by IEEE of primarily conference papers from 2009-2011 (McCook, 2018).
- But that does not explain their continued lack of proper classification years after retraction.
- PubMed and Web of Science have a small number of incorrect document types relative to the size of the databases; however, 50+ improperly indexed articles are still a concern for quality control.

**Conclusions**

- The poor indexing of retracted articles has implications for:
  - Bibliometric researchers who want to find all retracted publications.
  - Domain scientists who want to eliminate retracted research from their searches.
  - Domain scientists who want to identify the retraction status of a single item.
- Database providers should devote greater attention to quality control, even just correcting document types of those items beginning with “RETRACTED ARTICLE” or those retracted by IEEE would be an improvement.
- Researchers should consult multiple sources to confirm whether or not an article is retracted.

**Acknowledgments**

Alfred P. Sloan Foundation G-2020-12623, Retraction Watch for use of their data.

**References**