On Retraction Cascade? Citation Intention Analysis as a Quality Control Mechanism in Digital Libraries

Muhammad Usman(✉) and Wolf-Tilo Balke

Institute for Information Systems, TU Braunschweig, Braunschweig, Germany
{usman,balke}@ifis.cs.tu-bs.de

Abstract. The amount of information in digital libraries (DLs) has been experiencing rapid growth. With the intense competition for research breakthroughs, researchers often intentionally or unintentionally fail to adhere to scientific standards, leading to the retraction of scientific articles. When a paper gets retracted, all its citing articles have to be verified to ensure the overall correctness of the information in digital libraries. Since this subjective verification is extremely time and resource-consuming, we propose a triage process that focuses on papers that imply a dependence on retracted articles, thus requiring further reevaluation. This paper seeks to establish a systematic approach for identifying and scrutinizing scholarly works that draw upon retracted work by direct citations, thus emphasizing the importance of further evaluation within the scholarly discourse. Firstly, we categorized and identified the intention in the citation context using verbs with predicative complements and cue phrases. Secondly, we classified the citation intentions of the retracted articles into dependent (if the citing paper is based on or incorporates part of the cited retracted work) and non-dependent (if the citing article discusses, criticizes, or negates the cited work). Finally, we compared the existing state-of-the-art literature and found that our proposed triage process can aid in ensuring the integrity of scientific literature, thereby enhancing its quality.

Keywords: Retraction Analysis · Citation Intention · Digital Libraries

1 Introduction

Building new research results upon existing work is a central pillar of scientific progress. The existing quality control peer-review process serves as a solid and (to some degree) accepted foundation for developing new ideas. Moreover, it serves as a benchmark to assess some new ideas' plausibility and possible benefits, then allows discussing, reproducing, expanding, or challenging previous results, but also to contrast new ideas against the current state of the art. Supporting this discourse according to the FAIR principles[1] is a central responsibility for modern digital libraries. Here digital libraries,

[1] While the FAIR principles were originally designed for scientific data management and stewardship, their adaptation to scientific publications is quite straightforward, see FAIR Principles - GO FAIR (go-fair.org).

O. Alonso et al. (Eds.): TPDL 2023, LNCS 14241, pp. 117–131, 2023.
https://doi.org/10.1007/978-3-031-43849-3_11

on the one hand, act as classical knowledge providers to make scientific results findable and accessible. On the other hand, they also have to actively ensure that publications can quickly and safely be used and are bound to offer conflicting, inconsistent, and sometimes even contradictory content due to comprehensively representing scientific discourse. Yet over time, they will provide a rich and commonly accepted body of new insights building on and citing the original research.

Due to a growing number of instances, processes for quality control in digital libraries also need to reflect on how to deal with publications that suffer from scientific misconduct, e.g., plagiarism, fabrication, or falsification [1]. The retractions of publications are not caused by differing opinions, experimental results, or theories but by either an intention to deceive or at least gross negligence of scientific standards. The governance of the retraction process is currently limited to alerting the scientific community when some previously published article has been found to include either explicit misconduct, such as deliberately misleading claims or fabricated data, or other serious errors that render a study's results and conclusions unreliable or irreproducible. Indeed, the number of retracted articles is continuously increasing across fields. According to *Retraction Watch* (RW), the number of retracted studies increased by 800% between 2010 and 2020 compared to before 2010. As of April 2023, the *RW* database[2] lists over 43,000 retractions, including the reasons for these retractions.

However, what should a governance structure for handling retractions within a concise digital library need to consider? A retraction does not only affect the retracted paper but may also affect all papers citing the original research in the worst case, leading to a cascade of retractions. This paper focuses on the triage process to screen out the studies that need reconsideration because their scientific argument depends on the referenced retracted article. We conducted our experiments on over 1000 citing papers of retracted articles. Our experimental findings assist in determining the papers that are partially or entirely dependent on a cited retracted work, i.e., which need to be reevaluated to ensure whether the retraction of a cited paper does not change the findings of citing work.

2 Related Work

Retraction is a prolonged process requiring extensive discussion and investigation to raise serious concerns [2]. The objective of a retraction is to discredit the alleged article and alert the scientific community about its validity. In related work, we focus on citation intention analysis and citation behavior of retracted articles.

2.1 Citation Intention Analysis

A citation context is a concise summary of the concept described in the respective cited reference, consisting of one or multiple contiguous sentences. It provides evidence from scientific literature to support, explain, or build a hypothesis [3]. However, citations of the scientific article are of unequal importance depending on the intention of the citation [4]. The citation analysis offers insight into citations' qualitative and quantitative behavior.

[2] Retraction Watch Database (retractiondatabase.org).

More than 150 types of citation intention classes exist in the literature [8], and several studies have investigated citation intentions, e.g., [6, 7].

Jiang et al. performed citation intention analysis and made an intriguing observation. They found that the models with the best overall performance were not the best in per-class performance [9]. This suggests that citation intention classification remains a challenging task that requires further exploration and development of more specialized classification for individual categories. Te et al. (2022) investigated the specialized categories of citation intention (critical and non-critical), which is crucial for identifying potential errors and encouraging self-correction of scientific findings [10]. More recent works categorize the intention of citations as important, non-important [11, 12], and influential, non-influential [13]. Numerous ontologies are available to classify the intent of the citation, including FaBiO or CiTO [14]. The CiTO classifies citation intention into 41 empirical categories. However, if we had perfect ontology annotation of citation context, would that solve to cascade citations of retracted articles?

2.2 Citation Behavior of Retracted Articles

The period from publication to retraction takes up to three years on average. During this period and even after the retraction, the citations of retracted articles continued to increase [18]. Of course, it is permissible to cite a retracted article, provided the reader is made aware of the retraction. Yet, several authors have raised concerns about the frequent use of retracted articles without referencing the retraction notice [19].

Research on the citation behavior of retracted articles focused on quantitative aspects, such as citation growth and alt-metrics [20]. In 2021, Heibi and Peroni performed a citation analysis of Wakefield's retracted work [21], claiming an (in reality non-existent) association between vaccinations and autism [22]. They found that Wakefield's citations continued to increase after retraction, but most citations are for general discussions. Heibi and Peroni recently performed quantitative and qualitative analyses on the citations of retracted articles in humanities [23]. They observed no decline in the total number of citations following the retraction. However, a few citations' contexts expressed a negative sentiment. In addition, Heibi and Peroni observed that the health sciences have a higher level of awareness about citing a retracted publication than the humanities and social sciences.

In 2020, Fu and Schneider introduced a system to determine if an article's conclusion is based on cited work [24, 25]. Their approach showcased promising results; however, the system requires substantial contributions from human experts, making the process arduous. For instance, in their case study on citing non-reproducible code, the researchers formulated specific questions to assess the impact on the citing article. Despite its novelty, the subjective mining of arguments and the need for substantial input from experts deterred the implementation of the system. In contrast, this paper proposes a novel and automated approach to categorizing articles by identifying the dependency between citing and cited retracted articles.

3 Methodology

Identifying and flagging papers that reference and, to some degree, depending on claims from retracted articles remain a persistent challenge for state-of-the-art digital libraries. To address this, we primarily focused on the following research question: *To what extent can we reliably distinguish between the citations of the retracted articles that depend on it and those that do not?* Our process involves collecting, annotating citations, and developing a classification model to predict which articles require reevaluation (see Fig. 1).

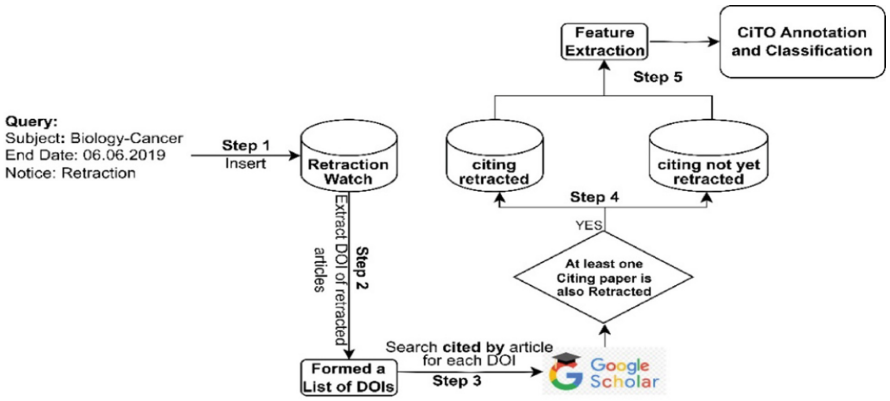

Fig. 1. We started with a collection of citations of retracted articles, identified the instances where both citing and cited articles are retracted, and then extracted their features. Secondly, we classified the citations based on the behavior of the citations into dependent and non-dependent.

3.1 Dataset Selection

To collect papers that cite the retracted articles, we first used *Retraction Watch*[3], a free, comprehensive online database, to form a list of retracted articles. We considered the retracted articles from the Biology-Cancer domain as a preliminary study. We then collected citing papers of retracted articles using *Google Scholar*. The existing datasets on citations of retracted articles suffer from limitations, such as incomplete coverage of all citations, potential biases resulting from analyzing different subsets of citing articles, and reliance on individual high-profile retraction case studies [24–27]. Given these limitations, considering various retraction cases, considering all their citations is of utmost importance to maintain a fair investigation.

Our selection process was rigorous and explicit. Our search endeavors yielded intriguing observations, revealing a recurring phenomenon wherein articles citing retracted publications undergo subsequent retractions. These compelling instances lend

[3] Retraction Watch Database (retractiondatabase.org).

substantial credence to our underlying hypothesis, highlighting the necessity for reassessment when referencing articles associated with retraction. Therefore, after careful consideration, we sought out retracted articles, ensuring that each article contained at least one instance where the citing article was also retracted. Given this condition, we found 28 articles and considered all of their subsequent, over 1000 citing articles, as a preliminary investigation. We extracted the title, authors, informed citation (If an author cited the retraction notice also or at least included the word "Retract", "Retraction" "Retracted" in reference or citation text), self-citations, citation context (one sentence before and after citation sentence), frequency of citation, citation section, journal, publication year, and Digital Object Identifier (DOI) for all articles.

3.2 Taxonomy Building

The citation intentions' definition often overlaps and has diffuse meaning. Furthermore, the annotator's mental models create a proliferation of competing perspectives that may conflict [16]. There are several ontologies, but we opted for the CiTO developed by SPAR because it describes a citation's nature *factually* and *rhetorically*. No other OWL ontology provides as extensive a set of citation properties as CiTO [15]. However, not all properties are equally adaptable. Some categories lead to overlapping meanings, and projects often use fragments of the CiTO distinctive properties, e.g., Journal of Cheminformatics [17]. To reduce the risk of getting the contradictory annotation, we categorize CiTO properties into five groups that cover all possible distinct intentions of referencing a scientific article; "*inconsistent*", "*discuss*", "*consistent*", "*use material*", and "*build-upon*" based on their given CiTO definition (see Table 1). For instance, we found that the CiTO functions of "*discusses*" and "*describes*" both relate to explaining the cited work, while "*confirms*" and "*cites as evidence*" are associated with showing similarities with the cited work. Therefore, by using rhetorical definitions, we grouped "*discusses*" or "*describes*" into "*discuss*" and "*confirms*" or "*cites as evidence*" into "*consistent*". We then divide the defined five groups into "*dependent"* and "*non-dependent"* based on the impact each

Table 1. Categorization of CiTO-Properties

Inconsistent	Discuss	Consistent	Use Material	Build-upon
deride ridicule refutes critiques disagree disputes corrects retracts	parodies, qualifies, credit, discusses, describes, reviews, Information, quotation, Related, metadata, authority, Assertion, replies, background,recommended documents, data source, compiles, excerpt from, links to, plagiarizes, solution, agrees with	obtain support cites as evidence, speculates on, confirms	uses { conclusions, data, method}	updates, extend

category can have on the citing article *(*see Table 2). On the one hand, articles that incorporate a part of a retracted work demonstrate dependence, and those that show consistencies with an unreliable source of information also require a close look. On the other hand, articles discussing, negating, contrasting, or criticizing the retracted work in citation context show the non-dependence, which is acceptable and does not harm scientific literature. The generalized decision rules for determining the dependency in the citation context are shown in Table 3.

Table 2. Definition and Example of Dependent and Non-Dependent Categorize

Intent Category and Definition			Example
Non-Dependent	Inconsistent Discuss	The citation provides context for the problem, concept, method, topic, and field relevance or credits, discusses, recommends, critiques, refutes, or disagrees with the work	It was found that supplementation with curcumin, a dominant component of Indian spice, could upregulate miRNA-200 and downregulate miRNA-21 [58]
Dependent	Consistent Use Material Build upon	The citation shows consistencies and similarities in results/ conclusion with the cited work or uses its method, data, or conclusion to build on it	The MTT assay was performed as described previously [19] These findings are in agreement with our observation that FoxM1 upregulates MMP-2 and MMP-9 in U2OS osteosarcoma cells and support the notion that MMP-2 and MMP-9 play a role in FoxM1-dependent tumor invasion [50]

Table 3. The generalized rules for determining the dependency of citation context

Analyze	Action
Does the citation context incorporate a part of retracted work?	Yes: Citation context is dependent No: Proceed to the next question
Does the citation context show consistency with the finding of retracted work?	Yes: Citation context is dependent No: proceed to the next question
Does the citation context discuss the retracted work?	Yes: Citation context is non-dependent No: Further assessment may be required

3.3 Citation Intention Annotation

To identify the dependency in the citation context, we utilized Stanford linguist Levin's instructions for *verbs with predicative complements* [29]. Due to their grammatical complementarity, these verbs are more naturally classified than nouns or prepositional phrases [30]. We extracted the verbs from the citation context with the help of the Stanford NLP toolkit, which implies dependency. However, the extracted verbs can also be part of the citation context with a non-dependency citation intent, where they are incorporated for some other purpose instead of referring to dependency on retracted work. Two university graduates with expertise in semantics, document analysis, and text classification conducted the annotation process. They observed verbs with predictive complements and identified cue phrases that may indicate the presence of dependency in the citation context. We also formed a list of example cue phrases from our dataset, indicating the dependency on cited work (see Table 4).

Table 4. The table shows the Verbs with Predicative Complements and Cue phrases that can imply dependency in the citation context.

Verbs	Cue Phrases
Incorporate, consider, consist, derive, depend, adopt, employ, produce, extend, confirm, prove, apply, induce, use, compare, add, suggest, base, support, relate, verify, ensure, promote, facilitate, indicate, accept, establish, reveal, obtain, agree, influence, similar, favor, yield, illustrate, encourage, compliment, convince, validate, evidence, evolve, inspire, emphasize, utilize	In line with [R], Consistent with [R], supported by a recent report [R], In agreement with [R], experiments performed as [R], Similar to [R] Procedures described in [R] Substantiated by recent evidence [R], Aligned with the findings in [R],

To further mitigate the likelihood of mistakes in the annotation process, we collaborated with an English language expert to address potential linguistic ambiguities. This collaborative effort enabled us to establish a 0.87 inter-agreement score. We discovered that most citations to retracted articles focused on discussing the referenced article, compared to only 11% employing the content of referenced article materialistically. We now perform statistical analysis and build a robust classifier to effectively discern the citations with dependency on the cited retracted article.

3.4 Citation Patterns of Retracted Articles

The retraction of scientific articles is a critical problem, as it can harm scientific progress and damage the reputation of the scientific community. Therefore, it is essential to understand how often retracted articles are cited, how researchers cite, and whether or not authors know about the retraction. Through an analysis of the various correlations outlined below that we observed in our study, it is possible to gain insights into how the scientific community responds to the issue of retracted articles. This can aid in

identifying areas where additional efforts are required to enhance awareness and prevent the dissemination of misinformation.

Pre- and Post-Retraction Citations: Modern digital libraries such as Google Scholar and PubMed emphasize retracted studies to inform readers about the credibility concerns present in literature. Despite this, the citations of retracted articles continued to rise. However, we observed that, after some time, the citations dropped sharply; only 17% came in the five years following the retraction, compared to 83% in the previous five years.

Informed Citation of Retracted Articles: It is acceptable to cite retracted work, provided the author is aware of the retraction. It is recommended to reference both the original publication and the retraction notice, as they offer digital object identifiers (DOIs). Despite the considerable efforts made by DLs to bring attention to the retraction, only 1% of the references after the retraction contained the term "retracted" within the reference. This finding implies that the author's lack of awareness regarding the retraction could be attributed to either negligence or insufficient dissemination of the retraction notice. This required the scientific community's attention to take further steps to prevent the propagation of misinformation.

Pre- and Post-Retraction Dependent Citations: Researchers cite an article in good faith in the pre-retraction period, as they are not aware of the retraction at that time. However, in post-retraction time, the lack of awareness and ignorance of authors about the retraction caused no significant difference in the number of dependent instances after the retraction. We found that 15% and 13% of citations in pre- and post-retraction times are dependent, respectively. Having dependent citations in pre-retraction time can be considered an honest mistake, but only cases of gross negligence can result in such an outcome during the period following retraction.

Dependent Self-Citations: It is common practice for scientists to do self-citations to expand their research. However, it could have severe consequences if the referenced literature is subsequently retracted. It was observed that a notable proportion of dependent citations consist of self-citations, and all of the self-citations dependent on the original work are present in the pre-retraction phase and are in good faith, in contrast to only one in the post-retraction phase. It raised concerns about the self-citations of retracted articles and required intention from the scientific community to evaluate such cases critically.

Frequency of Citation: The "frequency of citation," or the number of times a specific article is referred to within the body of a citing article, is considered a strength of the correlation between the citing and cited articles. We observed that 40% of the time, papers containing potentially dependent citations referred to a retracted article more than once.

4 Experimental Settings and Result

We utilized a dataset of citations manually extracted from scientific articles for experiments. The citation text contains markups to references in the bibliography, such as [12], (12), (Author et al.), which is not helpful for the classifiers that we aim to design

to identify dependencies in text. To remove citation references from the text, we employ a regular expression (\([^\)]*,[^\)]*\)). A few citation references also required manual omission, as there are different styles for referring to articles in the text, for example, a superscript type of reference. We used the NLTK library for preprocessing data for machine learning classifiers. The NLTK library contains "no", "nor", and "not" words as stop words, which are useful in the citation context, so we have excluded such words from the list that can carry the semantic meanings. In addition, to deal with words with multiple terms, we used bigram vectorization for classification with conventional machine learning algorithms. We conducted experiments using 10-fold cross-validation, with an 80–20 split for training and testing. We conducted experiments using 10-fold cross-validation. Our dataset contains imbalanced classes, so we focused on weighted precision, recall, and F1 scores as our evaluation metrics. Weighted precision, recall, and F1-score are vital evaluation metrics for imbalanced datasets. Weighted precision measures the accuracy of positive predictions, and Weighted recall gauges the model's ability to capture actual positive instances. Weighted F1-score combines precision and recall, providing a balanced overall performance assessment. Our experiments involved the use of several machine learning classifiers, including naive Bayes, logistic regression, and support vector machine, as well as deep learning models such as Long Short-Term Memory (LSTM) [31] and Bidirectional Encoder Representations from Transformers (BERT) [32]. By using multiple models and considering various performance metrics, we were able to gain a comprehensive understanding of the effectiveness of our approach.

In the experimental setting for LSTM, we utilized the Keras tokenizer. The tokenized sequences are prerequisites for data input into the model. We designed the neural network architecture with an embedding layer, an LSTM layer with dropout regularization, and a dense layer with a sigmoid activation function. We chose the LSTM layer due to its proven suitability for processing sequential data such as text. The dropout regularization was implemented to prevent overfitting, while the dense layer used a sigmoid activation function to assign a probability score to each class. In the experimental setting for BERT, we used a pre-trained BERT-base-uncased model. The model is fine-tuned for the "dependent" and "non-dependent" categories. We used the Adam optimizer [33] with binary cross-entropy as the loss function to optimize the model's performance for LSTM and BERT. We trained the models for ten epochs to promote comprehensive learning from the data and achieve improved classification performance; by hit and trial, we found that more epochs add no value.

Our experiments revealed that deep learning models, particularly the LSTM and BERT, outperformed conventional classifiers in categorizing dependent and non-dependent instances. The LSTM achieved the highest weighted precision of 0.94, followed by BERT with 0.90 (See Table 5). The LSTM has achieved better results than a BERT model, and for a small dataset, it gets trained faster than tuning the pre-trained counterparts [34], causing a better performance than BERT. The machine learning models SVM, logistic regression, and naive Bayes also yield promising results in classifying dependent and non-dependent instances and achieved weighted precision of 0.89, 0.87, and 0.81, respectively. Given that the data is not comprehensive for training, there is room for improvement. Ambiguities in language introduce the potential for false-positive and false-negative classifications. For instance, the citation context, "*The study by Johnson*

Table 5. Classification of "Dependent" and "Non-Dependent" Citation Intention

Classifier	Weighted-Precision	Weighted-Recall	Weighted-F1-score
SVM	0.89	0.85	0.87
LR	0.87	0.81	0.84
Naïve Bayes	0.81	0.80	0.80
LSTM	0.94	0.93	0.94
BERT	0.90	0.91	0.90

et al. provided valuable insights into the topic, which were further discussed in this paper." can be interpreted as both dependent and non-dependent. This ambiguity highlights the challenge of accurately determining the dependency status of citation contexts. Enhancing the training dataset with more diverse examples and incorporating additional contextual information is vital to disambiguate such cases and improve classification performance. The supplementary material utilized in this study is available for access[4].

The implications of our findings are significant in terms of identifying articles that cite and depend on retracted articles. Our classifiers provide the ability to flag such articles, ensuring the integrity of digital libraries. This capability allows us to effectively pinpoint articles that require further scrutiny, thereby preventing the potential dissemination of unreliable information. By considering our approach, researchers and library curators can play a crucial role in maintaining the quality and trustworthiness of the literature. This, in turn, fosters an environment conducive to reliable and accurate scientific advancements.

5 Qualitative Comparison with a Manual Approach

In 2022, Addepalli et al. [25] tested a keystone framework [24] to find the dependency of citing articles on Wakefield's retracted article [21]. It was published in 1998, and after years of discussion, Wakefield's work was partially retracted in 2004 and fully retracted in 2010 [5]. According to *Retraction Watch*[5], the reasons for its retraction are *data* fabrication and *result* manipulation. We have applied our approach to the same set of citing articles for comparison as used by Addepalli et al.

In Addepalli et al.'s work [25], two annotators performed the annotation based on the flow chart. One annotator marked two instances as *dependent*, nine as *up to professional*, and the second marked two as *dependent* and seven as *up to professional*. Both marked the remaining citing papers as *independent* of Wakefield's work. However, they agreed in only one instance of *up to professional* before the discussion. This significant pre-discussion divergence, caused by the conflicting understanding of the annotation rules or a flow chart's disputed meaning, leads to different results. We identified five articles that dependently cited Wakefield's retracted article. In two of those

[4] https://github.com/Conferences2023/TPDL.

[5] Retraction Watch Database (retractiondatabase.org).

instances, we found an agreement with Addepalli et al. The article with dependency either cites Wakefield's work to show consistency with their result or incorporates the methodology or data. Overall, we found an agreement of above 90%, where most citing articles independently cited Wakefield's work. Table 6 compares contradicting predictions about dependency or non-dependency with the reason for our different judgments. Our approach has several benefits over the existing approach. Firstly, our approach is not limited to case studies, where an explicit argument must be defined to uncover dependency, making it highly adaptable and applicable. Secondly, we recognize citations' vital role in representing rhetorical relations and information flow in linking scientific articles within digital libraries and show how information in the articles is interconnected. Furthermore, the extension of our approach holds the potential to contribute to quality control in digital libraries significantly. By incorporating our methodology into modern digital library systems, we can introduce a layer of scrutiny to identify and stop the spread of misinformation.

Additionally, we conducted a metaphorical comparison with *Scite*[6]. It categorized the citations into *supporting*, *monitoring*, and *contrasting*. The *supporting* category represents citations with identical results. The *monitoring* encompasses citations with discussions, and the *contrasting* category includes citations that present differing opinions. During our analysis, we observed that two instances we marked as dependent were also categorized as supporting in Scite. Furthermore, one of the dependent instances from Addepalli et al.'s study is categorized as independent by our approach and monitoring by Scite. However, it is essential to note that Scite does not currently encompass all the citations available. For instance, Wakefield's article is cited in over 4000 articles listed in *Google Scholar*, whereas Scite only displays less than 2500 citations.

Table 6. Comparative Analysis of Diverging Judgments and Contributing Reason

PMID	Existing work [3]	Our Approach	Reason
12142948	up to professional	dependent	incorporate data
15622451	independent	dependent	based on
19917212	independent	dependent	consistent findings
16003130	independent	dependent	incorporate data
12773694	dependent	independent	Discussion
15526989	dependent	independent	Discussion
15031638	dependent	independent	Discussion

6 Discussion and Conclusion

Scientific articles are not stand-alone entities but are interconnected by citations. Researchers cite existing articles to make an argument for their new findings. When an article gets retracted, citing articles whose conclusion depends on it must also be

[6] Scite: see how research has been cited.

reconsidered. Despite increasing citations of retracted articles, modern digital libraries do not flag such cases. In this paper, we effectively analyzed and classified over 1,000 citations of retracted articles based on the intention in the citation context into dependent and non-dependent. We can infer that the reliability of findings is questioned when an article indicates the dependency on a fallacious source of information. In an ideal world, we can eliminate this problem if all citations are explicitly annotated with the intention of its citation so that when a cited paper gets retraction, its subsequent citing paper that depends on its fallacious claims also gets retraction.

Moreover, we uncovered some instances where citing and cited papers were retracted, despite the citing paper having no dependency on the retracted work. Such retraction cases are identified based on subjective evaluation from journal editors or third-party investigators. Retraction is a critical task, and the existing state of art process takes up to three years on average to retract scientific articles. This opens the discussion of how close we are to having a system capable of retracting a paper and underscores the need for a more robust retracting system. We identified four cases of citations of retracted articles based on citation intention (see Fig. 2).

	Dependent	Non- Dependent
Retracted	Case 1: 2.9%	Case 2: 6.0%
Non-Retracted Yet	Case 3: 7.7%	Case 4: 83.5%

Fig. 2. The correlation between the citation intention and the retraction

Case I: When a cited article gets retraction due to errors or misconduct, which raises questions about the accuracy and reliability of the citing paper. In such cases, it is essential to consider the impact of these errors on the citing paper. Depending on the extent to which the citing paper depends on the retracted work, it may be necessary to retract the citing paper as well, to prevent the spread of misinformation. Case 4: Retraction is a procedure carried out by publishers and editors to indicate that an article is invalid due to misconduct or errors that undermine the credibility of the findings. There is no harm in citing retracted work if the researcher is aware of its retraction status and employs it solely for background information or to discuss the scientific problem at hand. It is imperative to acknowledge that retracted literature can impede scientific progress and compromise the integrity of subsequent research. However, it remains plausible that the aforementioned article gets retracted due to potential misrepresentation of its methodology, data, or results based on an arbitrary investigation, as discussed in [28] (Case 2). We identified instances to triage the process to screen out the studies that implied dependency on cited retracted articles that are not retracted yet (Case 3). Therefore, we must thoroughly investigate such instances before concluding about the subsequent retraction.

7 Limitation and Future Work

Although our study provides valuable insights for citations of retracted articles, we identified a potential shortcoming in our current approach. Our current approach relied on explicit indicators in the citation context that implied dependency on the cited work. It

can be problematic when the paper implicitly implies dependence, which could result in papers depending on retracted work without any indication in the citation context being overlooked. In future work, we will explore more sophisticated methods for identifying implicit indicators and consider the full text to categorize articles that require reevaluation. Moreover, recognizing the need for a more comprehensive understanding of retractions in science, we plan to expand our scope and aim to formulate a comprehensive dataset and train language models to use it across fields to identify articles with dependencies on unreliable sources of information.

Acknowledgments. Supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): PubPharm – the Specialized Information Service for Pharmacy (Gepris 267140244).

References

1. Resnik, D.B.: From baltimore to bell labs: reflections on two decades of debate about scientific misconduct. Account Res. **10**, 123–135 (2003). https://doi.org/10.1080/08989620300508
2. Bar-Ilan, J., Halevi, G.: Temporal characteristics of retracted articles. Scientometrics **116**(3), 1771–1783 (2018). https://doi.org/10.1007/s11192-018-2802-y
3. Hernández-Alvarez, M., Gomez, J.M.: Survey about citation context analysis: tasks, techniques, and resources. Nat. Lang. Eng. **22**(3), 327–349 (2016). https://doi.org/10.1017/S1351324915000388
4. Aljohani, N.R., Fayoumi, A., Hassan, S.U.: An in-text citation classification predictive model for a scholarly search system. Scientometrics **126**(7), 5509–5529 (2021). https://doi.org/10.1007/s11192-021-03986-z
5. The Editors of The Lancet Retraction—Ileal-Lymphoid-Nodular Hyperplasia, Non-Specific Colitis, and Pervasive Developmental Disorder in Children. Lancet **375**, 445 (2010). https://doi.org/10.1016/S0140-6736(10)60175-4
6. Cohan, A., Ammar, W., van Zuylen, M., Cady, F.: Structural scaffolds for citation intent classification in scientific publications. In: Proceedings of NAACL-HLT, pp. 3586–3596 (2019). https://doi.org/10.18653/v1/N19-1361
7. Bakhti, K., Niu, Z., Yousif, A., Nyamawe, A.S.: Citation function classification based on ontologies and convolutional neural networks. In: Learning Technology for Education Challenges: 7th International Workshop, LTEC 2018, Žilina, Slovakia, August 6–10, 2018, Proceedings 7, pp. 105–115. Springer International Publishing (2018). https://doi.org/10.1007/978-3-319-95522-3_10
8. Ihsan, I., Qadir, M.A.: CCRO: citation's context & reasons ontology. IEEE Access **7**, 30423–30436 (2019). https://doi.org/10.1109/ACCESS.2019.2903450
9. Jiang, X., Cai, C., Fan, W., Liu, T., Chen, J.: Contextualised modelling for effective citation function classification. In: Proceedings of the 2022 6th International Conference on Natural Language Processing and Information Retrieval, pp. 93-103 (2022). https://doi.org/10.1145/3582768.3582769
10. Te, S., Barhoumi, A., Lentschat, M., Bordignon, F., Labbé, C., Portet, F.: Citation context classification: critical vs. non-critical. In: Proceedings of the Third Workshop on Scholarly Document Processing, pp. 49–53 (2022)
11. Valenzuela, M., Ha, V., Etzioni, O.: Identifying meaningful citations. In: AAAI Workshop: Scholarly Big Data, vol. 15, p. 13 (2015)

12. Wang, M., Zhang, J., Jiao, S., Zhang, X., Zhu, N., Chen, G.: Important citation identification by exploiting the syntactic and contextual information of citations. Scientometrics **125**, 2109–2129 (2020). https://doi.org/10.1007/s11192-020-03677-1
13. Zhu, X., Turney, P., Lemire, D., Vellino, A.: Measuring academic influence: not all citations are equal. J. Assoc. Inf. Sci. Technol. **66**, 408–427 (2015). https://doi.org/10.1002/asi.23179
14. Peroni, S., Shotton, D.: FaBiO and CiTO: ontologies for describing bibliographic resources and citations. J. Web Semant. **17**, 33–43 (2012). https://doi.org/10.1016/j.websem.2012.08.001
15. Mayernik, M.S., Phillips, J., Nienhouse, E.: Linking publications and data: challenges, trends, and opportunities. D-Lib Mag. **22**(5/6), 11 (2016). https://doi.org/10.1045/may2016-mayernik
16. Ciancarini, P., Iorio, A.D., Nuzzolese, A.G., Peroni, S., Vitali, F.: Evaluating citation functions in CiTO: cognitive issues. In: European Semantic Web Conference, pp. 580–594. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07443-6_39
17. Willighagen, E.: Two years of explicit CiTO annotations. J. Cheminformatics **15**(1), 14 (2023). https://doi.org/10.1186/s13321-023-00683-2
18. Hsiao, T.-K., Schneider, J.: Continued use of retracted papers: temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine. Quant. Sci. Stud. **2**, 1144–1169 (2021). https://doi.org/10.1162/qss_a_00155
19. Bar-Ilan, J., Halevi, G.: Post retraction citations in context: a case study—Scientometrics **113**(1), 547–565 (2017). https://doi.org/10.1007/s11192-017-2242-0
20. Feng, L., Yuan, J., Yang, L.: An observation framework for retracted publications in multiple dimensions. Scientometrics **125**, 1445–1457 (2020). https://doi.org/10.1007/s11192-020-03702-3
21. Wakefield, A., et al.: RETRACTED: ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet **351**, 637–641 (1998). https://doi.org/10.1016/S0140-6736(97)11096-0
22. Heibi, I., Peroni, S.: A qualitative and quantitative analysis of open citations to retracted articles: the Wakefield 1998 et al.'s case. Scientometrics **126**(10), 8433–8470 (2021). https://doi.org/10.1007/s11192-021-04097-5
23. Heibi, I., Peroni, S.: A quantitative and qualitative open citation analysis of retracted articles in the humanities Quant. Sci. Stud. 1–46 (2022).https://doi.org/10.1162/qss_a_00222
24. Fu, Y., Schneider, J.: Towards knowledge maintenance in scientific digital libraries with the keystone framework. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020. ACM: New York, NY, USA, August 2020; pp. 217–226 (2020). https://doi.org/10.1145/3383583.3398514
25. Addepalli, A., Subin, K.A., Schneider, J.: Testing the keystone framework by analyzing positive citations to Wakefield's 1998 paper. In: International Conference on Information, pp. 79–88. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-96957-8_9
26. Schneider, J., Ye, D., Hill, A.M., Whitehorn, A.S.: Continued post-retraction citation of a fraudulent clinical trial report, 11 years after it was retracted for falsifying data. Scientometrics **125**, 2877–2913 (2020). https://doi.org/10.1007/s11192-020-03631-1
27. Hsiao, T.K., Schneider, J.: Continued use of retracted papers: temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine. Quant. Sci. Stud. **2**(4), 1144–1169 (2021). https://doi.org/10.1162/qss_a_00155
28. Williams, P., Wager, E.: Exploring why and how journal editors retract articles: findings from a qualitative study. Sci. Eng. Ethics **19**(1), 1–11 (2013). https://doi.org/10.1007/s11948-011-9292-0
29. Levin, B.: English verb classes and alternations: a preliminary investigation. University of Chicago Press (1993)

30. Kipper, K., Korhonen, A., Ryant, N., Palmer, M.: A large-scale classification of English verbs. Lang. Resour. Eval. **42**, 21–40 (2008). https://doi.org/10.1007/s10579-007-9048-2
31. Sherstinsky, A.: Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D **404**, 132306 (2020). https://doi.org/10.1016/j.physd.2019.132306
32. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018). https://doi.org/10.48550/arXiv.1810.04805
33. Zhang, Z.: Improved Adam optimizer for deep neural networks. In: 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), pp. 1–2. IEEE (2018). https://doi.org/10.1109/IWQoS.2018.8624183
34. Ezen-Can, A.: A comparison of LSTM and BERT for small corpus. arXiv preprint arXiv:2009.05451 (2020). https://doi.org/10.48550/arXiv.2009.05451