Implementation of Academic Misconduct Prevention Functions in Semantic Publishing of Scientific Journals: Postprint

Authors: Wen Li, enumerate

Date: 2023-10-08T00:00:00+00:00

Abstract

To provide reference for the development of semantic publishing in scientific journals and the prevention of academic misconduct, this study examines the implementation of academic misconduct prevention in the semantic publishing process of scientific journals based on their development trends and the characteristics of semantic publishing. In terms of academic misconduct prevention, semantic publishing introduces two major changes to the scientific journal publishing workflow: in peer review, shifting from sole duplicate detection to innovation prediction; and in published content, transitioning from single-article publishing to linked data publishing. The prevention function against academic misconduct in scientific journals can be realized through the intelligence of detection software, the interlinking of published content, the precision of content delivery, and the decentralization of the publishing process. However, conflicts persist in the semantic publishing of scientific journals, including those between open data sharing and publishers' existing interests, as well as conflicts arising from the joint application of semantic technology and blockchain technology. This paper proposes several recommendations to address these issues.

Full Text

Realizing Academic Misconduct Prevention Functions in Semantic Publishing of Scientific Journals

Abstract: To provide references for the development of semantic publishing in scientific journals and the prevention of academic misconduct, this study examines the implementation of academic misconduct prevention during the semantic publishing process of scientific journals, based on publishing trends and the characteristics of semantic publishing. In terms of misconduct prevention, semantic publishing introduces two major changes to the publishing workflow:

shifting from simple duplication detection to innovation prediction in peer review, and moving from single-article publishing to linked data publishing in content dissemination. Through intelligent detection software, linked content publishing, precise content delivery, and decentralized publishing processes, scientific journals can effectively realize their academic misconduct prevention functions. However, conflicts remain between data openness and publishers' existing interests, as well as between the joint application of semantic and blockchain technologies. This paper offers recommendations to address these issues.

Keywords: Semantic publishing; Academic misconduct; Plagiarism detection; Linked publishing; Blockchain

Classification Code: G237
Document Code: A
Article ID: 1671-0134(2019)10-025-03
DOI: 10.19483/j.cnki.11-4653/n.2019.10.005

Authors: Wen Li, Luo Lie

In recent years, academic misconduct has exhibited new patterns. With easier access to information, lower costs of plagiarism, increasingly diverse forms, and stronger concealment, such misconduct is gradually becoming normalized. Its detrimental impact on the academic environment has drawn growing attention. As vital platforms for exchanging scientific achievements, scientific journals represent the final line of defense for intervening in and preventing academic misconduct, playing a crucial role in purifying the academic environment. Currently, misconduct prevention in journal publishing workflows relies primarily on peer review, plagiarism detection systems, and requiring authors to sign "integrity pledges" during the review stage. However, both peer review and emerging detection software have limitations, while integrity pledges have largely become formalities with minimal effectiveness, driving academic misconduct toward greater concealment and detection difficulty [1].

Existing research on misconduct prevention during the publishing stage mainly focuses on patterns of misconduct and preventive measures. For instance, Xu Shiyong et al. summarized and analyzed author misconduct cases encountered by the journal *Silk* in recent years and proposed targeted countermeasures [2]; Wu Ning analyzed evolving characteristics of misconduct in scientific journals and suggested response strategies [3]; and Wang Zijun et al. discussed how to identify misconduct during manuscript revision [4]. Studies indicate that since all approaches have loopholes, the most critical pathway for preventing misconduct in scientific publishing remains strengthening journal team building and improving editor competence, as detection often depends on editors' sense of responsibility and professional expertise, requiring full utilization of their subjective initiative [5-7]. This inevitably leads to occasional oversights.

As an emerging publishing paradigm, research on semantic publishing has concentrated on technical implementation within publishing workflows, with scarce attention to its role in preventing academic misconduct. Based on the applica-

tion of semantic technology in scientific journals, this study elaborates in detail and from multiple perspectives how semantic publishing can technically prevent academic misconduct, aiming to provide references for averting such behavior in the development of semantic publishing and enabling scientific journals to serve as important guardians of the academic ecosystem and impartial instruments of academic evaluation.

1. Changes in Scientific Journal Publishing Workflow for Academic Misconduct Prevention

1.1 From Simple Plagiarism Detection to Innovation Prediction

Due to the time lag between initial plagiarism detection and online publication, editors may need to conduct multiple similarity checks before final publication. True semantic publishing begins during manuscript creation. Through seamless integration between the article and detection systems, every modification triggers automatic real-time similarity detection, with automatic alerts when thresholds are exceeded, greatly facilitating editorial work. At the review stage, semantic publishing not only prevents misconduct in peer review and innovation assessment but also helps editors evaluate a manuscript' s research background, disciplinary progress, and application prospects, moving beyond simple similarity detection.

The application of semantic technology facilitates establishing a truly collaborative platform across the entire network with minimal effort, maximizing intellectual property protection and academic misconduct detection. Evolving from simple duplication detection to networked, real-time checks and then to innovation assessment, semantic technology plays a vital role in early misconduct detection and screening out low-level repetitive research.

1.2 From Single-Article Publishing to Linked Data Publishing

Traditional journal space limitations prevent authors from providing detailed data analysis processes, facilitating data falsification and fabrication. To improve research reproducibility, researchers increasingly expect journals to provide original research data. Submitting supporting data during submission and publishing it upon acceptance has become a basic requirement for many leading scientific journals such as *Science*, *Nature*, and *Cell*, yet traditional publishing' s limited space cannot satisfy this demand [8].

In 2014, *Scientific Data* launched by Nature Publishing Group and the "Global Change Scientific Research Data Publishing System" established by the Institute of Geographic Sciences and Natural Resources Research (CAS) and the Geographical Society of China both implemented linked publishing models pairing entity data with data papers. Currently, over 30 domestic journals including *Acta Geographica Sinica* have reached consensus on publishing linked data papers, with some journals already requiring submission of supporting data. For

example, *Chinese Journal of Health Management* has gradually required original research submissions to provide corresponding raw materials such as original data, results, scales, intervention methods, and questionnaires since 2016. However, due to limitations of traditional and current data publishing, these materials cannot fully realize their potential.

2. Implementation Pathways for Academic Misconduct Prevention in Semantic Publishing

2.1 Intelligent Detection Software

Plagiarism detection software can identify some academic misconduct, yet its vulnerabilities and limitations render misconduct more concealed and difficult to detect. Semantic technology not only transforms publishing content but also enables more intelligent and accurate detection. Currently, one semantic technique employed in detection software involves word frequency statistics, derived from vector space retrieval models related to information retrieval technology. This method first counts word occurrences in each document, then converts word frequencies into spatial feature vectors according to specified rules, and finally calculates similarity by measuring vector distances to achieve plagiarism detection [9].

In recent years, CNKI' s Research Integrity Management System has effectively integrated content-based information fingerprinting with semantic analysis technology, ensuring detection accuracy while providing strong anti-interference capabilities. The system supports detection of various transformation forms including rewriting, restructuring, and translation, achieving semantic-level content detection. It can perform similarity detection on special objects like figures and tables by combining titles, context, and content [10], and conduct intelligent information classification based on specific concepts, viewpoints, and conclusions. As semantic technology matures in detection systems, similarity analysis incorporating adjustable semantic granularity uses sentences as the basic analysis unit for rapid and accurate text comparison. This powerful detection capability will intensify academic misconduct screening, exposing even general low-level repetition and leaving common evasion tactics nowhere to hide.

2.2 Linked Content Publishing

By fragmenting paper content and linking it with original data information, the complete research process can be reproduced, making data sources clearer, article quality more trustworthy, and enabling reanalysis of paper content to objectively evaluate its value and impact on scientific development. Replication is an effective means to verify the authenticity and scientific validity of published content; by repeating operations using the paper' s data and methods and comparing experimental results with original data, the authenticity, credibility, and reliability of research outcomes can be judged. Utilizing semantic technology for linked data publishing can fully display experimental processes, provide a

foundation for replication, and increase the likelihood of detecting misconduct. Sharing these resources through semantic publishing also enables more data users to actively participate in distinguishing paper authenticity from multiple perspectives. Although data falsification and fabrication cannot be completely eliminated, the costs required for such fraud itself···

2.3 Precise Content Delivery

Massive amounts of online information are indiscriminately delivered to different users. General proactive delivery not only causes user anxiety but is also easily ignored. Tailor-made precise delivery can solve this problem. By analyzing user preference profiles, semantic publishing can intelligently deliver published content to experts and scholars in relevant fields, enabling timely discovery of ghost authorship and unauthorized authorship cases. Due to professional limitations, general research circles have their bounded scope, and semantic publishing' s precise intelligent delivery increases the exposure probability of plagiarism to target audiences, making such situations nowhere to hide.

2.4 Decentralized Publishing Process

Current copyright protection models can no longer meet contemporary needs, and a simple "integrity pledge" merely constitutes a moral constraint. Semantic publishing practices adopt FAIR principles for data release, which overlap significantly with decentralization principles: independent data storage from service providers, interoperability, accessibility, and the ability to freely express and process network trust [11]. Semantic publishing naturally possesses decentralized characteristics that prevent arbitrary article tampering. True semantic publishing begins semantic analysis and processing during the manuscript creation stage [11]; if semantic publishing starts from researchers'experimental data recording, the probability of subsequent data tampering and copyright disputes will substantially decrease.

3. Obstacles to Realizing Misconduct Prevention Functions

3.1 Conflict Between Data Openness and Publishers' Interests

Simple duplication detection alone cannot fully leverage semantic technology' s advantages. Academic misconduct detection requires the most comprehensive possible full-text comparison resource database. Currently, CNKI' s Research Integrity Management System can compare against documents in CNKI's "China Knowledge Resource Integrated Database" and partially open internet resources, but cannot access other databases. Semantic publishing and data openness are mutually beneficial, enabling networked collaboration in misconduct detection, yet this conflicts with the interests of database publishers.

3.2 Conflict in Joint Application of Semantic and Blockchain Technologies

Semantic publishing represents limited decentralization. By combining with other technologies to design trustworthy semantic publications, complete decentralization can be achieved, such as through blockchain technology. Blockchain is a collectively maintained distributed database formed cryptographically, where data blocks are chained together, nodes mutually supervise each other, traditional third parties are removed from the workflow, and information is interconnected yet tamper-proof, ensuring high security [12]. Blockchain enables copyright registration for internal data blocks, and its consensus mechanism makes copyright transactions transparent [13]. Although blockchain' s application in copyright remains exploratory, it represents an effective intellectual property protection method. Wang Zhong et al. proposed a conceptual verification system for the joint application of both technologies in agricultural product quality and safety traceability [14]. Applying blockchain technology on top of semantic publishing, establishing consensus mechanisms through smart contracts, can secure data access and rights attribution, adding trust bonds and data security guarantees to scientific paper publishing. Semantic technology makes information usable, while blockchain ensures privacy from illegal acquisition and data security; a certain conflict exists between them. Although integrating blockchain and semantic technologies proves highly effective for preventing academic misconduct, its feasibility requires further validation.

4. Conclusions and Recommendations

Semantic publishing is a digital content publishing model characterized by elastic scalability, dynamic allocation, and resource sharing. It spawns numerous new publishing information services and drives overall transformation of the publishing landscape. Fully leveraging semantic technology' s powerful analytical capabilities can improve plagiarism detection accuracy, enhance identification rates of academic misconduct from multiple dimensions through semantic publishing, and prevent misconduct via technical means. This facilitates more scientific control and management of scientific journal publishing workflows, establishes protective barriers for journal copyright, ensures sustainable and healthy journal development, and fosters a sound academic environment.

Against the backdrop of open science, the trend toward open sharing of publishing data is evident. In semantic publishing, a model combining open sharing with partial paid usage can be implemented. For misconduct detection, linking interests with various publishers and combining comprehensive free services with personalized paid services can enhance their motivation to share data, thereby achieving networked collaborative detection. Currently, semantic content organization has become a major trend in scientific journal development in the big data era. However, to enable widespread authors and editors to skillfully apply semantic technology in writing and editing, user-friendly software interfaces

must be developed. Journal publishers should emphasize professional talent cultivation and knowledge updating for existing staff. Well-resourced institutions may consider recruiting specialized technical personnel, while established journal clusters can adopt remote sharing of professional intellectual resources.

Funding: Youth Fund of China University Journals Association (CUJS-QN-2018-029) and General Project of Hunan Provincial Department of Education "Application of Semantic Technology in University Scientific Journal Development" (14C1192)

Author Affiliation: Journal Press, Central South University of Forestry and Technology

References:

[1] Wang Wenfu. Strategies for Mining Hidden Academic Misconduct in Journal Papers in the Internet Era [J]. Chinese Journal of Scientific and Technical Periodicals, 2016, 27(7): 677-682.
[2] Xu Shiyong, Ye Jing, Kang Feng, et al. Phenomena, Causes and Preventive Measures of Academic Misconduct in Journals [J]. Acta Editologica, 2019, 31(4): 411-414.
[3] Wu Ning. Evolution of Academic Misconduct in Scientific Journals and Countermeasures [J]. Acta Editologica, 2019, 31(4): 369-371, 376.
[4] Wang Zijun, Zhao Liqin. Research on Identifying Academic Misconduct During Manuscript Revision [J]. Public Communication of Science & Technology, 2019, 11(13): 146-147.
[5] Lu Yixin. Analysis of Hidden Repetition Phenomena in Academic Misconduct Detection [J]. Chinese Journal of Scientific and Technical Periodicals, 2015, 26(6): 552-555.
[6] Peng Fang, Jin Jianhua, Dong Yanping. Analysis of Causes of Peer Review Fraud and Preventive Measures [J]. Acta Editologica, 2018, 30(3): 240-243.
[7] Xu Haili. Academic Misconduct and Its Preventive Measures [J]. Chinese Journal of Scientific and Technical Periodicals, 2015, 26(6): 545-551.
[8] Kong Lihua, Xi Yan, Lang Yangqin, et al. Research on Peer Review Methods for Scientific Data in Data Journals [J]. Acta Editologica, 2019, 31(3): 262-266.
[9] Na Na. Research on Plagiarism Detection Technology and Its Application in Information Verification [D]. Tianjin: Nankai University, 2008.
[10] Sun Xiongyong. Analysis Report on the Application of Academic Misconduct Literature Detection Systems in Academic Journals [N]. China Press, Publication, Radio, Film and Television News, 2016-12-30(8).
[11] Xu Lei. Several Research Hotspots in Semantic Publishing Worthy of Attention [J]. Publishing Science, 2018, 26(2): 16-22.
[12] Sun Jing, Guan Qingshan, Duan Lihui, et al. Current Status of Media Convergence Publishing in Scientific Journals and Reflections on Digital Copyright Protection and Construction [J]. Chinese Journal of Scientific and Technical Periodicals, 2018, 29(8): 813-821.
[13] Niu Min. Research on Digital Copyright Management Models Based on Blockchain Technology [D]. Beijing: Beijing Institute of Graphic Communica-

tion, 2017.
[14] Wang Zhong, Liang Zaoqing, Zheng Yelu, et al. Application of the Intersection of Semantic Web Technology and Blockchain Technology in the Field of Agricultural Product Quality and Safety Traceability [J]. Agricultural Library and Information, 2019, 31(1): 60-68.

Note: Figure translations are in progress. See original paper for figures.