

Full Length Research Paper

Intricacies of utilizing Turnitin tool in agricultural extension content writing in Nigeria

Emmanuel Odinaka Okoroma^{1*}, Ogbonnaya Aja Okoro², Christiana Ihuoma Agomuo²,
Comfort Chigozie Godson-Ibeji² and Nsongurua Dianabasi Inyang¹

¹Department of Agricultural Economics and Extension, Faculty of Agriculture, University of Uyo, Uyo, Akwa Ibom State, Nigeria.

²Department of Agricultural Extension, School of Agriculture and Agricultural Technology, Federal University of Technology Owerri, Imo State, Nigeria.

Received 29 July, 2023; Accepted 25 August, 2023

The study analyzed how Turnitin misjudges semantics used in agricultural extension writing in Nigeria. The paper applied thematic content analysis on 30 selected extension contents from Nigerian sources. Codes, percentage count and line graph were used in analysing data. The results show the inability of Turnitin to recognize agricultural extension semantics, such as, “The study was designed to...”, and “The result shows that...” which were flagged most with 81.0 and 75.0% occurrence, respectively. The text similarity was highest in the methodology section of the work, followed by the literature review section for empirical papers while text similarity peaked in the main body of the paper for non-empirical papers. The paper concludes that Turnitin algorithm is a text matching tool that cannot recognize semantics used in extension in Nigeria. Manual review and/or recalculation of Turnitin text similarity index were recommended.

Key words: Extension content, plagiarism, integrity, intellectual property, text similarity Turnitin.

INTRODUCTION

The advancement in electronic communication and machine learning is changing the way agricultural information and technologies are being disseminated. Notably, the new paradigm is increasingly pushing the process of agricultural information creation into machine learning while of content dissemination is increasingly being moved to and from online repositories (Meo and Talha, 2019). In machine learning, the use of text similarity tools like Turnitin in extension has become an integral means of promoting academic integrity and protecting intellectual properties. On the other part, online channels like institutional websites, organizational

websites and online databases, social media, and prerecorded audio-visual content have offered effective means of transmission and consumption of extension content (Ekerete et al., 2021).

Extension content connotes agricultural information, written articles, field data, materials, transcripts, documentaries and other forms of intellectual properties of extension professionals transmitted electronically to extension audience and for extension purpose (Meuschke et al., 2019; Orisakwe and Okoroma, 2020). The emphasis on intellectual property implies that the exclusive right of ownership of such contents is given by

*Corresponding author. E-mail: oserv55@yahoo.com.

law to the designated owners (National Universities Commission (NUC), 2021). The clause of “designated owner” implies ownership by registration, protection, and licensing with relevant institutions, agencies of the government and organizations that propagate, protect, transmit and consume intellectual property. It is on the basis of this law that academic writing protocols which criminalize the act of impropriety like plagiarism is enforced across the 99 private Universities, 45 Federal Universities and 52 State Universities in Nigeria (National Universities Commission (NUC), 2021).

Plagiarism entails acts that undermine academic integrity such as, intentionally or mistakenly personalizing other people’s work rather than giving credit to owners, fabricating field data, hoarding useful research information, falsifying academic achievements to attract credit, among other falsehood (Meuschke et al., 2018a; Sulistiani and Karnalim, 2019; Bensalem, 2020; Prakoso et al., 2021). The pressure to publish, poor training in writing protocols, ignorance, misunderstanding, weak and defective academic misconduct policies highly promote acts of plagiarism (Meo and Talha, 2019). As a deterrent, different institutions in Nigeria have deployed Turnitin as a plagiarism detection tool with specific benchmarks.

Unfortunately, Turnitin and other text similarity detectors like iThenticate, PlagScan, Plagtracker, Unicheckwork on the mechanism of text similarity, which merely matches text with uploaded contents in global repositories, but lack the capability to establish where and when the text similarity has amounted to plagiarism (Meuschke et al., 2019; Sulistiani and Karnalim, 2019). Hence, the exclusive use of Turnitin as a sole determinant of plagiarism as practised in many institutions in Nigeria misrepresents what constitutes plagiarism and equally misapplies Turnitin as a plagiarism detector rather than as a text marching tool.

Numerous comprehensive studies have highlighted the complexities and limitations in the application of Turnitin as a text similarity detection tool. Notably, research conducted by Meuschke et al. (2018b), Prasetya et al. (2018), Meuschke et al. (2019), and Meo and Talha (2019) has extensively explored these intricacies. However, there appears to be a gap in the literature regarding the demonstration of how semantic terms, acronyms, and specialized expressions employed in the context of agricultural extension in Nigeria could be both identified and potentially misconstrued as instances of text similarity. Emerhirhi et al. (2020) argued that semantics peculiar to extension practice in Nigeria if treated in a general context could be misrepresented. Also, previous works have failed to show the trajectory of Turnitin text matching for Nigeria-based extension contents as well as how the text similarity index can be manually reviewed and calculated for a more accurate originality report.

Against the backdrop of the information gap from previous studies on Turnitin use in extension in Nigeria,

the study specifically:

- i) Examined how Turnitin manipulates semantic terms, acronyms and expressions used in agricultural extension in Nigeria during text matching;
- ii) Analyzed Turnitin text similarity detection trajectory on selected Nigerian Extension contents;
- iii) Demonstrated how text similarity index is manually calculated and/or recalculated.

METHODOLOGY

The paper adopted content analysis on selected extension contents from Nigerian sources. Nigeria is located at 9.0820°N and 8.6753° E with a total land area of 923,768 Km². A 5-stage content analysis technique as modeled by Gaur and Kumar (2018) was used in establishing the content inclusion and exclusion criteria. First, five major research repositories/search engine, namely: Google scholar, Scopus, Web of Science, Research gate, Ajol, were purposively searched due to their large repository of peer extension publication. Second, “agricultural extension in Nigeria”, “extension development programmes”, “methodology”, “introduction”, “theoretical review”, “conceptual review”, “objectives of the study”, “themes”, were among the combined keywords used in the search terms. Third, the search period covered extension works published between 2000 to 2022. Fourthly, the search scope was limited to extension articles specific to Nigeria or Nigerian location. In the fifth stage, the following steps in Table 1 were taken to include and exclude irrelevant articles.

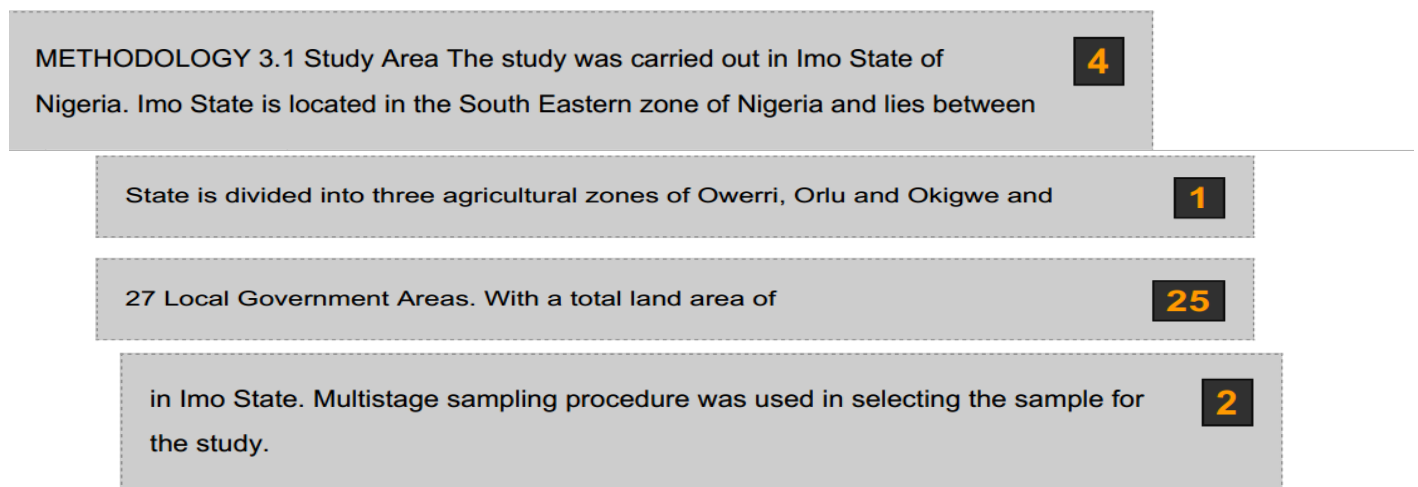
The 30 extension contents used in the study comprised of 15 empirical extension papers and 15 position articles and were sourced from the Journal of Agricultural Extension (JAE), the Postgraduate repository of the University of Uyo (UniUyo) and the Agric4Africa blog for meeting the inclusion criteria. Precisely, for the empirical papers, 10 journal papers were drawn from JAE and 10 Master’s Degree were drawn from Uni Uyo, respectively. While, for non-empirical contents, 5 journal position papers and 5 blog articles were selected on the basis of meeting the inclusion criteria. The purposive use of empirical papers and position articles was to achieve sample inclusiveness and multiple text pattern (Krisppendorff, 2018).

Furthermore, the selected contents were subjected to Turnitin test to determine their text similarity pattern. The fourth stage involved thematic coding of Turnitin text similarity reports for the 30 selected contents. For the empirical contents, code numbers 1 denoted flagged Acronym, 2 denoted flagged expression, 3 denoted flagged term, while 4 denoted flagged subtheme under abstract, introduction, literature, methodology, result and discussion, conclusion and recommendations, while for non-empirical contents code numbers 1 denoted flagged Acronym, 2 denoted flagged expression, 3 denoted flagged term, while 4 denoted flagged subtheme under introduction, main body, conclusion and recommendations. In the final stage, the frequencies of the assigned codes were further aggregated into a frequency distribution. Data analysis was carried out using frequency distribution, percentage count, line graph and bar chart.

Precisely, the text similarity detection on semantic expressions/acronyms used in agricultural extension content writing in Nigeria was analyzed using percentage count carried out on Turnitin report of the selected 30 extension contents. Turnitin Text Similarity Detection Trajectory on selected Nigerian Extension Papers was addressed using line graph generated from the frequency distribution of Turnitin reports of the extension contents, while Turnitin text similarity pattern for selected non-empirical Nigerian extension contents was achieved using manual text

Table 1. Systematic inclusion and exclusion of articles for content analysis.

Items	Number
Extension articles on Nigeria identified from initial database search	1777
Articles considered relevant to extension semantics by hand-searching the Postgraduate repository of the University of Uyo (UniUyo)	12
Duplicated articles removed	788
Total number of articles screened	1.001
Articles excluded due to low occurrence of extension semantics in the work	663
Articles selected with the full text	109
Articles considered eligible based on the set inclusion and exclusion criteria	87
Articles excluded due to missing extension semantics in the introduction, literature review, methodology and discussion, as well as not being peer-reviewed	57
Total number of articles used for the study	30

**Figure 1.** Pictorial showing Turnitin non-recognition of semantics used in agricultural extension in Nigeria. Source: Turnitin Reports Extract (2022).

similarity index calculator performed on the Turnitin report. Figure 1 is an instance illustrating the inability of Turnitin algorithm to recognize semantic terms, expressions and acronyms, as it often flags them in text matching thereby erroneously inflating the text similarity index which is in turn used exclusively by many institutions in deciding for plagiarism (Meo and Talha, 2019).

RESULTS AND DISCUSSION

Text similarity detection on semantics

Table 2 presents the distribution of Turnitin Text Similarity Detection showing a list of semantic expressions/acronyms used in Agricultural Extension Content writing in Nigeria and in the inability of the software to recognize them as semantics that need no citation or to be flagged as text similarity. According to the result of the Turnitin tests carried out on 30 extension contents, the semantic expression, “The result shows that...” was flagged most

by the software (90.0%), and closely followed by the expression, “The study was designed to...” (81.0%). Some other expressions successively flagged included, ‘Hence, a gap exists in knowledge’, ‘To describe the socioeconomic characteristic of ...’, ‘...State is located on longitude’, ‘Major economic activities in the State...’, ‘The study used primary data collected with the aid of a questionnaire...’, ‘The study used multistage sampling technique...’ (70.0%).

The high percentage occurrence of similarity detection strongly shows the inability of Turnitin to recognize semantics used in Nigerian extension. This discovery aligns with the previous claims made by Meuschke et al. (2018a), Prasetya et al. (2018), Meuschke et al. (2019), Meo and Talha (2019), Bensalem (2020), and Wang and Dong (2020) that Turnitin lacks the ability to comprehend the contextual nuances in which specific expressions are employed to convey distinct meanings. In Nigeria, the aforementioned expressions provide stereotype format

Table 2. Text similarity detection on semantic expressions/acronyms used in agricultural extension content writing in Nigeria.

Semantic Expression/Acronyms	Number of cases subjected to Turnitin test	Number of times detected as text similarity	% occurrence
The study was designed to...	22	18	81.0
*Hence, a gap exists in knowledge	20	14	70.0
The following research questions are considered pertinent...	20	09	45.0
The specific objectives of the study include....	20	12	60.0
To describe the socioeconomic characteristic of ...	20	14	70.0
To determine the factors influencing...	20	10	50.0
Conceptual framework	20	07	35.0
Theoretical Framework	20	09	45.0
Empirical framework	20	07	35.0
The literature review is presented under the following subthemes...	20	09	45.0
Agricultural Development Program (ADP)	25	17	68.0
ORiver Basin Development Authority (RBDA)	25	09	36.0
Operation Feed the Nation (OFN)	25	09	35.0
National Fadama Development program (NFDP)	25	10	40.0
Anchor Borrowers Program (ABP)	25	09	36.0
Study area	20	08	40.0
The study was carried out in...	20	11	55.0
State is located on longitude	20	14	70.0
State is bordered on the...	20	11	55.0
Major economic activities in the State...	20	14	70.0
The study used primary data collected with the aid of a questionnaire...	20	14	70.0
The study used multistage sampling technique...	20	14	70.0
Data analysis was carried out using descriptive and inferential statistical tools	20	10	50
The study used content validity in validating the research instrument	20	09	45.0
Split half method was used in establishing the reliability of the research instrument	20	06	30.0
Table...is the distribution of respondents by socioeconomic characteristics	20	12	60.0
The result shows that...	20	18	90
The study concludes that ...	25	11	44
Based on the findings of the study the following recommendations are made	25	11	44

and means of conveying information in extension, as such, should not be considered as text similarity or plagiarized statements (Okoroma et al., 2021).

However, regardless of the above deficiency, Turnitin at moment is the most reliable text similarity detection tool used in Nigeria. Consequently, there arises a necessity to employ

alternative methods for validating Turnitin's outcomes in the context of extension. Furthermore, when extension content writers endeavor to rephrase or break down statements highlighted by

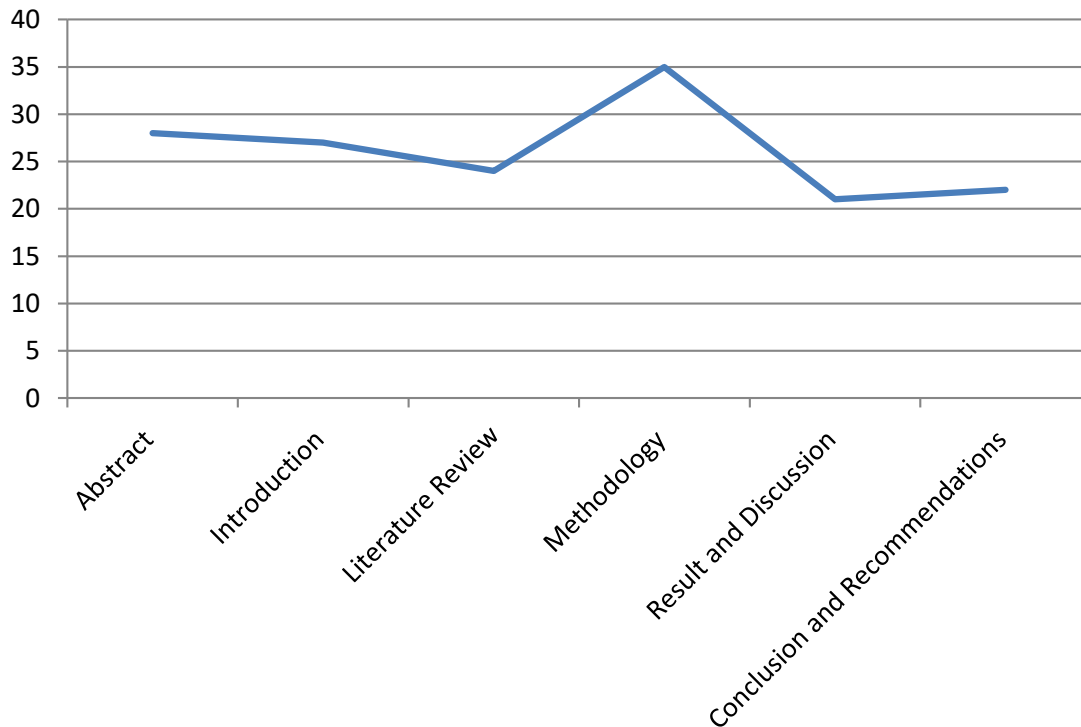


Figure 2. Turnitin text similarity pattern for selected empirical Nigerian extension contents.

Turnitin, there is a risk of distorting the original intended meaning.

Turnitin text similarity detection trajectory on selected Nigerian extension papers

Figure 2 is a graphical representation of Turnitin text similarity detection trajectory carried out on empirical extension contents using the mean occurrence. The result indicates that text-similarity of the extension contents was highest in the methodology section of the work, followed by the literature review section for empirical papers. Figure 3 on the other hand, presents the text similarity trajectory of non-empirical (position) extension papers. The result of the content analysis reveals that text similarity detection by Turnitin peaks in the main body of the paper.

Drawing from the earlier finding of Okoroma et al. (2021), the above result is due to the stereotypical nature of extension paper writing. For instance, extension research methodology involves replicating procedures, techniques and locations. Another reason according to Foltýnek et al. (2019) is due to the inability of text similarity algorithm to recognize semantic text undermines the efficiency of automated text similarity detection. Little wonder, Turnitin often flags common terms, expressions, and acronyms used in extension practice and communication in Nigeria as plagiarized text. In the case of the literature review, the high detection of text similarity

commonly results from the replication of stereotype subthemes and common language.

Implicitly, to avoid being flagged on text similarity, authors need to replace or paraphrase stereotype terms and expressions used in extension content in Nigeria. While this approach appears feasible, it is likely to distort the technical meaning and nomenclature of such irreplaceable terms and expressions (Meo and Talha, 2019). For instance, it is not possible to replace stereotype subheadings like methodology, literature review; names of locations in Nigeria; terms and acronyms peculiar to extension in Nigeria such as ADP, NALDA, OFN, NAFPP, etc, without distorting the technical meaning.

Manual calculation of text similarity index

The need to demonstrate manual calculation of text similarity index is sequel to the semantic issues associated with the use of Turnitin in agricultural extension content in Nigeria, as well as its primary use as a plagiarism detection tool. To enhance the aforementioned purpose text similarity index should be recalculated as shown below. Formula for manual for Text Similarity Index (TSI):

$$\text{TSI} = \frac{\text{Total number of words highlighted as text similarity}}{\text{Total word count of the paper}} \times 100$$

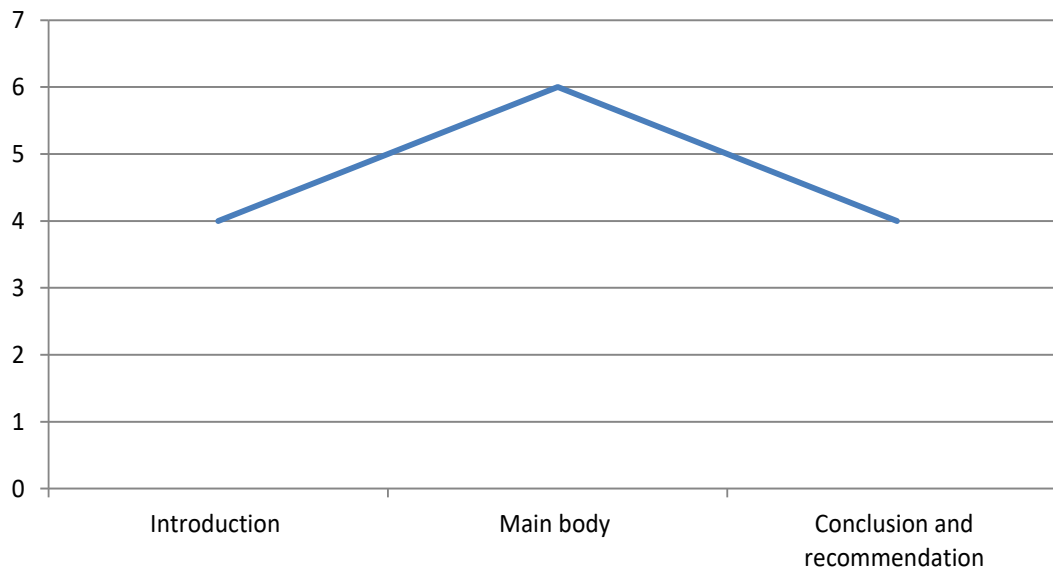


Figure 3. Turnitin text similarity pattern for selected non-empirical Nigerian extension contents.

Let us apply the above formula to Figure 1, assuming it is one paper/document, the text similarity index is realized thus:

$$\begin{aligned} \text{Total number of words highlighted as text similarity} &= 95 \\ \text{Total word count of the paper/document} &= 126 \\ \text{TSI} &= \frac{95}{126} \times 100 \\ &= 75\% \end{aligned}$$

The example above presents a simplistic way of calculating text similarity index. Secondly, it highlights the margin of error that could result from preclusion of semantic terms in text similarity index. In this particular instance of Figure 1, the Turnitin text similarity index is defective and cannot be used as a reliable basis in judging for academic misconduct like plagiarism as opined by Foltýnek et al. (2019).

On the other hand, recalculating Turnitin TSI manually in order to exclude semantics can be tedious, especially for large volume paper. It involves the following steps:

- i) Step 1: Review the text similarity index report to identify and count the number of semantic terms, acronyms and expressions across all highlighted text similarity in the report.
- ii) Step 2: Find the percentage of the number of semantic words to the total word count of the papers. The total word count is found at the front page of the Turnitin report.
- iii) Step 3: Subtract the calculated percentage value of the semantics from the text similarity index score, the difference is taken as the actual TSI value.

For example, using Figure 1 where the Total word count = 126; TSI = 75.0%. Let us say after reviewing

the TSI report 44 words were considered as semantics. The new TSI will be recalculated thus:

- 1) Find the percentage of 44 to 126 = $\frac{44 \times 100}{126} = 35.0\%$
- 2) Subtract 35.0% from 75.0 = $75\% - 35\% = 40\%$

The above has shown in a simplistic manner how a text similarity index is recalculated by considering semantics, and reduced from 75.0 to 40.0%. Although this technique may appear tedious, it is helpful in addressing lingering issues undermining the use of Turnitin for enhancing research integrity. Across several institutions in Nigeria many Postgraduate Theses and Dissertation have been suspended from proceeding on account of high text similarity index literarily considered as plagiarism. It is important to stress at this point that contrary to the practice in many institutions in Nigeria, the use of Turnitin is helpful in evading plagiarism and improving research integrity, but does not in itself constitute a tool for plagiarism detection (Meuschke et al., 2018a; Vysotska et al., 2018; Meo and Talha, 2019).

CONCLUSION AND RECOMMENDATIONS

Turnitin is a text matching tool which does not recognize peculiar terms, acronyms and expressions used in extension practice in Nigeria as semantics that should not be factored into the text similarity index. Text similarity occurs more in the methodology section, followed by the literature review section for extension empirical paper. Text similarity for non empirical papers is peak in the main body of the paper. Manual recalculation of text

similarity index suffices in extension when the Turnitin report is fraught with semantic mis-judgments that significantly put the reliability of the report into question.

1) Manual review of Turnitin report should be encouraged by institutions and users of extension intellectual properties. This is important in ensuring that semantic terms, expressions and acronyms used in extension practice in Nigeria have not been seen substantially flagged as text similarity, thereby inflating the text similarity index value with pseudo words that should be ignored. There have been cases of manually recalculated Turnitin reports that found 20 of 35% text similarity index value composed of semantic expressions.

2) Institutions and stakeholders of extension should ensure that Turnitin is not used as a plagiarism detection tool, rather as a text matching tool. Implicitly, its test report should not be seen as a sacrosanct verdict on the integrity of content. In some institutions in Nigeria Turnitin test has either been suspended or jettisoned due to semantic issues surrounding its reliability. There should be an academic integrity review committee that validates every text similarity index score before it is acted upon. They will have the prerogative to recommend for manual recalculation when necessary.

3) The challenge of semantics in the use of Turnitin requires extension experts in Nigeria to focus attention towards developing text similarity algorithm that recognizes extension semantics. This will offer a more efficient way of addressing the issue rather than the manual review that is tedious and time consuming. Part of the process entails expanding repositories of extension semantics through increased uploading of extension contents in global open and closed access repositories. Such as, by providing internet websites and connectivity for faculties, departments and individuals involved extension content creation through federal government digitalization programme. By so doing more Nigerian extension contents find their way online thereby expanding online database of Nigerian extension.

4) Across institutions in Nigeria emphasis on Turnitin application is usually inclined to portray it as a plagiarism detection tool.

This can be accomplished by providing internet websites and connectivity for faculties, departments and individuals involved in extension content creation through federal government digitalization programmes.

There is need to effectively acquaint students and teachers with the intricacies of using Turnitin. Not only will such efforts enrich the knowledge and skills of students and teachers, but also raise compliance with academic propriety.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

REFERENCES

- Bensalem I (2020). "Intrinsic Plagiarism Detection: A Survey". Plagiarism Detection: A focus on the Intrinsic Approach and the Evaluation in the Arabic Language (PhD thesis). Constantine 2 University. DOI:10.13140/RG.2.2.25727.84641.
- Ekerete BI, Okoroma EO, Onwuka D (2021). Social Media Utilization Dynamics by Agricultural Research Students in Akwa Ibom State, Nigeria. Proceedings of the Annual Conference of Freshwater Biological Association of Nigeria (FBAN) held University of Uyo, Uyo Akwa Ibom State, Nigeria September 12th -16th.
- Emerhirhi E, Nnadi FN, Okoroma EO (2020). Analysis of village extension workers' capability in producing electronic media for extension education in Imo State, Nigeria. Middle East Journal of Applied Sciences 10(4) DOI:10.36632/mejas/2020.10.4
- Foltynek T, Meuschke N, Gipp B (2019). Academic plagiarism detection: A systematic literature review. ACM Computing Surveys 52(6):1-42. doi:10.1145/3345317
- Gaur A, Kumar M (2018). A systematic approach to conducting review studies: An assessment of content analysis in 25 years of IB Research. Journal of World Business 53(2):280-289. Available at www.sciencedirect.com
- Krisppenddorff K (2018). Content analysis: An introduction to its methodology. California, USA. Sage Publication.
- Meo SA., Talha M (2019). Turnitin: Is it a text matching or plagiarism detection tool? Saudi Journal of Anaesthesia 13(1): 548. Available at www.ncbi.nlm.nih.gov
- Meuschke N, Gondek C, Seebacher D, Breitingner C, Keim D, Gipp B (2018a). An adaptive image-based plagiarism detection approach. Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries pp. 131-140.
- Meuschke N, Stange V, Schubotz M, Gipp B (2018b). HyPlag: a hybrid approach to academic plagiarism detection. The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval pp. 1321-1324.
- Meuschke N, Stange V, Schubotz M, Kramer M, Gipp B (2019). Improving academic plagiarism detection for STEM documents by analyzing mathematical content and citations. 2019 ACM/IEEE Joint Conference on Digital Libraries pp. 120-129. Available at www.ieeexplore.ieee.org
- National Universities Commission (NUC) (2021). Monday Bulletin. Available at www.nuc.edu.ng
- Okoroma EO, Ekerete BI, Aligbe-Kikanme T, Nwokorobia C, Okoma C (2021). Implications of deploying plagiarism detection software in Agricultural Extension Research in Nigeria. Proceedings of the Annual Conference of Freshwater Biological Association of Nigeria (FBAN) held University of Uyo, Uyo Akwa Ibom State.
- Orisakwe L, Okoroma EO (2020). Effects of indigenous communication methods on agricultural innovation dissemination among rural women farmers in Imo State, Nigeria. Journal of Community and Communication Research 5(2). Available at <https://jccr.sccdr.org.ng>
- Prakoso DW, Abdi A, Amrit C (2021). Short text similarity measurement methods: a review. Soft Computing 25(6):4699-4723. Available at www.link.springer.com
- Prasetya DD, Wibawa DP, Hirashima T (2018). The performance of text similarity algorithms. International Journal of Advances in Intelligent Informatic 4(1):63-69. Available at www.link.springer.com
- Sulistiani L, Karnalim O (2019). ES-Plag: Efficient and sensitive source code plagiarism detection tool for academic environment. Computer Applications in Engineering Education 27(1):166-182. Available at www.onlinelibrary.wiley.com
- Vysotska V, Burov Y, Lytvyn V, Demchuk A (2018). Defining author's style for plagiarism detection in Academic Environment. IEEE Second International Conference on Data Stream Mining and Processing pp. 128-133
- Wang J, Dong Y (2020). Measurement of text similarity: a survey. Information 11(9):42. Available at www.aclanthology.org