

Comparing the similarity index across iThenticate, Ouriginal, and Turnitin plagiarism detection software

MD. HAMIDUR RAHMAN^{1*}, MUHAMMAD SHAHIDUL ISLAM², SYED MURTAZA HUSSAIN ANDRABI³, J. P. SHARMA⁴, MD. NASIM REZA⁵

^{1,5}Department of Physical Education and Sports Science, Jashore University of Science and Technology, Jashore, BANGLADESH

²Office of Physical Education, Shahjalal University of Science and Technology, Sylhet, BANGLADESH

³Department of Physical Education and Sports Sciences, University of Delhi, Delhi, INDIA

⁴Indira Gandhi Institute of Physical Education and Sports Sciences, University of Delhi, Delhi, INDIA

Published online: February 29, 2024

(Accepted for publication February 15, 2024)

DOI:10.7752/jpes.2024.02051

Abstract:

In the current information technology era, plagiarism is a significant and critical issue in research. Plagiarism detection tools are essential in identifying instances of plagiarism. This study compared the similarity index generated by three leading plagiarism detection software platforms: iThenticate, Ouriginal, and Turnitin. Ten original documents (N = 10) were selected for analysis across the three software programs. The process involved first analyzing all documents with Ouriginal, then checking the same documents, followed by iThenticate, and Turnitin. These software programs generated originality reports detailing the number of matching sources, similar word counts, and an overall similarity index as a percentage. To detect notable differences within the dataset, a one-way ANOVA and a Tukey (HSD) post-hoc analysis were conducted. The threshold for statistical significance was established at $p < 0.05$. Statistical analysis revealed that while there was a significant variance in the similarity index across the tools iThenticate, Ouriginal, and Turnitin ($F(2, 27) = 5.436, p = .010$), there were no notable differences in the sources they matched ($F(2, 27) = 1.289, p = .292$). This suggests that the plagiarism detection capabilities may vary significantly among these tools, but the sources they identify as matches are largely consistent. However, the average values indicated that Turnitin had the highest mean similarity detection followed by iThenticate, and then Ouriginal. In this study, evaluating the similarity index can help verify the effectiveness of anti-plagiarism tools and safeguard researchers against committing plagiarism.

Key Words: iThenticate, Ouriginal, Turnitin, plagiarism software, similarity index

Introduction

Plagiarism, often defined as the act of stealing or literary theft, is a significant challenge in academic and research settings (Aronson, 2007; Khadilkar, 2018; Bhattacharya, 2016). It involves the unauthorized use of another person's intellectual work or thoughts, falsely claiming them as one's own, without giving due credit (Joy & Luck, 1999; McIntosh, 2013). Plagiarism involves not only deliberate acts like mosaic plagiarism, idea plagiarism, text plagiarism, and self-plagiarism but also unintentional one's such as improper paraphrasing or quoting, all of which entail presenting someone else's work as one's own (Luksanaprukksa & Millhouse, 2016; Helgesson & Eriksson, 2015; Joob & Wiwanitkit, 2018).

Plagiarism typically falls into two primary categories: textual plagiarism and conceptual plagiarism (Debnath, 2016). This dishonest behavior not only discredits the original creator's work but also damages the credibility of the academic domain (Harris, 2001). The implications of plagiarism extend beyond academic boundaries. Plagiarism in academia has risen due to easy online access to information, confusion over what counts as plagiarism, the pressure to publish, competitive settings, and poor knowledge of citation methods (Scanlon, 2003; Scanlon & Neumann, 2002; Martin, 2005; Usta & Koçak, 2021). This issue undermines researcher credibility, has legal consequences, and can lead to the withdrawal of published work, harming both institutional and personal reputations (Green, 2012).

To avoid plagiarism by strictly following academic writing rules and ethical standards in research papers, authors must efficiently rewrite, paraphrase, reference, and edit their work (Muşat et al., 2023; Andronic et al., 2022). In addition, research institutions and professional bodies should offer essential resources and motivate researchers to produce and publish academic papers featuring novel ideas and content (Narayan, 2023). In academic publishing, the acceptable level of text similarity differs. Generally, under 10% similarity is often acceptable (Garg & Nagpal, 2023), while some editors only allow up to 5% (Peh & Arokiasamy, 2008), and over 20% might prompt additional scrutiny (Swaan, 2010). Sometimes over 10% can lead to manuscript rejection by journal editors and academic institutions (Mahian et al., 2017; Kadam, 2018).

Plagiarism detection software, available as an online-based application, comes in free and commercial versions (Smart & Gaston, 2019; Memon & Mavrinac, 2020). Commercial options may involve subscriptions, contracts with academic institutions, or purchasable tokens for a set number of pages or words. Free versions typically offer limited page or word checks at no cost (Condurache & Bolboacă, 2022; Nkotagu, 2022). Free plagiarism tools typically check manuscripts using a limited web database, with specifics found on their homepage (Adithan & Surendiran, 2018). iThenticate, Ouriginal, and Turnitin are paid plagiarism detection tools commonly used by universities, research entities, publishers, and individuals on a credit basis to detect plagiarism. iParadigms LLC launched the online plagiarism detection service Turnitin in 1997 and later introduced iThenticate in 2004 (Meo & Talha, 2019).

Prio Infocenter launched URKUND in 2000, which was later rebranded as PI/Ouriginal in 2020 and subsequently acquired by Turnitin in 2021 (Ouriginal, 2023). iThenticate offers manuscript screening for 1,500 leading publishers worldwide, utilizing Crossref's extensive scholarly database. It scans through a comprehensive collection that encompasses 97% of the most-cited journals, over 89 million academic resources, more than 200 million open-access materials, and about 99.3 billion web pages, reviewing around 14 million documents annually (iThenticate, 2023). Meanwhile, Ouriginal, and Turnitin, provides advanced plagiarism detection by integrating text-matching with style analysis. Its extensive database includes around 47 billion web pages, 190 million academic papers from 97% of the top publications, and 1.9 billion student submissions (Ouriginal, 2023; Turnitin, 2023).

A similarity index percentage is a common standard for detecting plagiarism (Chung et al., 2017). Plagiarism software provides a “similarity index” or “similarity” which measures the extent to which a manuscript matches content from its archives (Manley, 2013). This process involves analyzing the manuscript against previously published material and stored data to quantify textual similarities (Adithan & Surendiran, 2018).

For example, if a document has 90 words similar to another source and the total word count of the document is 3000, calculating the similarity index involves multiplying the number of similar words by 100 and then dividing by the total word count. In this case, the similarity index would be 3%. In the context of a similarity index, matching sources refer to comprehensive references that exhibit similarity to primary sources, such as internet resources, published works, student papers, crossrefs, URLs, and other relevant sources. Primary matching sources are organized numerically in different colors and each is assigned a unique percentage indicating its similarity.

Undoubtedly, by evaluating the similarity index, researcher can effectively evaluate the accuracy of anti-plagiarism tools and also protect their intellectual work against plagiarism. This process involves comparing study content with existing materials to identify any overlaps, ensuring that work is original and free of unintended duplication. This proactive approach not only verifies the effectiveness of plagiarism detection software but also helps maintain the integrity of work. For this reason, every scholar should have access to reliable anti-plagiarism software to assist in expressing their original ideas effectively. Therefore, the present study aims to compare the similarity indexes generated by the iThenticate, Ouriginal, and Turnitin plagiarism detection software.

Material & Methods

Sample

Ten original documents (N = 10) were chosen to be analyzed for similarity index. These ten documents comprise two original research papers, four scholarly assignments, and four academic Power Point presentations. All documents were the researchers' own and unpublished English-language work. Paid plagiarism detection tools iThenticate, Ouriginal, and Turnitin were utilized to assess the similarity of ten documents. The process involved first analyzing the documents with Ouriginal, followed by iThenticate, and finally, Turnitin.

These software programs generated final originality reports, presenting the similarity index as percentage values. The analysis intentionally omitted references, bibliographies, and any directly quoted material from the documents. For each documents, the final report were to be saved in PDF format. Finally, this study compared the similarity index between iThenticate, Ouriginal, and Turnitin plagiarism software.

Statistical analysis

The process involved calculating descriptive statistics for each software tool, which included determining the mean, standard deviation, standard error of the mean, and the range spanning from the minimum to the maximum value. To identify significant variations in the dataset, a one-way ANOVA and a Tukey (HSD) post-hoc test were applied, with a significance threshold set at 0.05.

The presentation of mean values within homogeneous (Levene's test) subsets further supports the assumption of a normal distribution in the data.

Results

Table 1 Descriptive statistics

Variables	Tools	N	Mean	Std. Dev.	Std. Error	Minimum	Maximum
Similarity Index	iThenticate	10	8.30	4.83	1.53	3.00	17.00
	Ouriginal	10	5.10	3.51	1.11	0.00	8.00
	Turnitin	10	12.10	5.67	1.79	3.00	22.00
Matching Sources	iThenticate	10	6.30	6.88	2.18	2.00	25.00
	Ouriginal	10	3.70	3.09	.98	0.00	9.00
	Turnitin	10	8.00	7.23	2.29	2.00	27.00

Table 1 indicates that the mean and standard deviation of the similarity index for iThenticate, Ouriginal, and Turnitin plagiarism detection software were 8.30 ± 4.83 , 5.10 ± 3.51 , and 12.10 ± 5.67 , respectively. Additionally, the mean and standard deviation for matching sources were 6.30 ± 6.88 , 3.70 ± 3.09 , and 8.00 ± 7.23 for each software's, respectively.

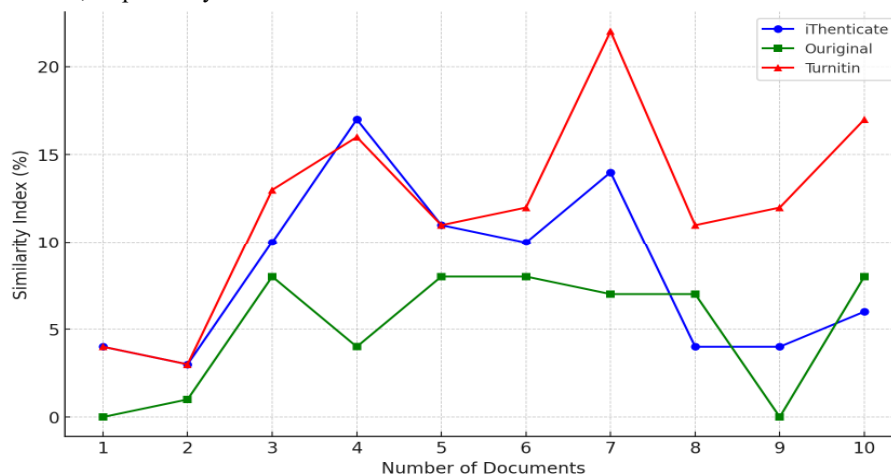


Fig. 1 Similarity index (%) for iThenticate, Ouriginal, and Turnitin plagiarism detection software

Table 2 One-way ANOVA

Variables	Groups	Sum of Squares	df	Mean Square	F-ratio	p-value
Similarity Index	Between Groups	245.60	2	122.80	5.436	.010
	Within Groups	609.90	27	22.59		
Matching Sources	Between Groups	93.80	2	46.90	1.289	.292
	Within Groups	982.20	27	36.38		

*. Significant at 0.05 level

Table 2 shows a significant difference in similarity index between groups ($F(2, 27) = 5.436, p = .010$). The finding indicates that there was a statistically significant difference in the similarity index among the plagiarism detection tools iThenticate, Ouriginal, and Turnitin. In contrast, matching sources did not show a significant difference ($F(2, 27) = 1.289, p = .292$). The statistical analysis revealed that there were no notable differences in the matching sources identified by the various software programs.

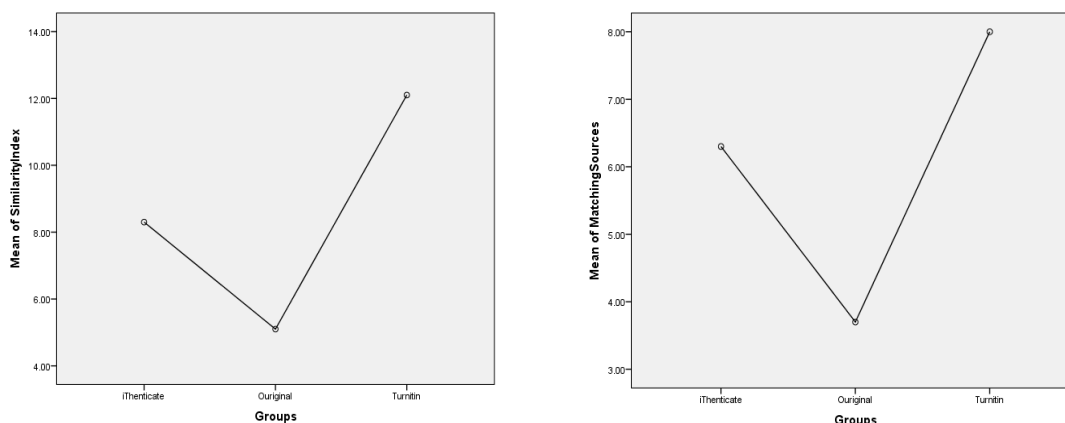


Fig. 2 Mean plots of similarity index % and matching sources in three tools

Table 3 Tukey (HSD) test

Variables	Tools	Groups	Mean Diff.	Std. Error	Sig. Level
Similarity Index	iThenticate	Ouriginal	3.20	2.13	.304
		Turnitin	-3.80	2.13	.193
	Ouriginal	iThenticate	-3.20	2.13	.304
		Turnitin	-7.00*	2.13	.008
	Turnitin	iThenticate	3.80	2.13	.193
		Ouriginal	7.00*	2.13	.008

*. The mean difference is significant at the 0.05 level

The Tukey Honest Significant Difference (HSD) post-hoc test (Table 3) indicated that the Turnitin similarity index was significantly different from that of Ouriginal plagiarism software ($p = .008$). However, no significant differences were found between iThenticate and Ouriginal ($p = .304$), or between iThenticate and Turnitin ($p = .193$).

Discussion

This analysis revealed notable distinctions in the similarity index of the iThenticate, Ouriginal, and Turnitin plagiarism detection tools. Specifically, Turnitin's similarity index differed significantly from Ouriginal's, while there were no marked differences between the index of iThenticate and Ouriginal or between iThenticate and Turnitin. Furthermore, in terms of average mean values, Turnitin indicates the highest similarity index, followed by iThenticate and then Ouriginal. Researcher Patra and Kirtania (2023) analysis of 20 themes across 83 paragraphs were conducted using three different plagiarism detection tools: iThenticate, Urkund (now known as Ouriginal), and Turnitin. The results showed similarity percentages for each software: iThenticate (7%), Ouriginal (7%), and Turnitin (13%). When examining the documents through iThenticate, Ouriginal (Urkund), and Turnitin for potential plagiarism, the results indicate a notably low level of similarity. The similarity index percentages reported by these plagiarism detection tools are iThenticate 12%, Ouriginal (Urkund) 1%, and Turnitin 5%, respectively (Kirtania, 2023). Baskaran et al. (2019) research examined 77 articles using iThenticate and Turnitin for similarity assessment. On average, Turnitin reported 8.66% similarity, while iThenticate showed 6.99%. Statistically, Turnitin's mean similarity index was significantly higher ($p < .0001$) than that of iThenticate. A doctoral thesis was analyzed by two anti-plagiarism tools, Ouriginal and Turnitin, yielding similarity indices of 6% and 0%, respectively, despite analyzing the same thesis text (Kale, 2019). In a study by Turnitin, which examined 310 papers from SCOPUS-indexed journals across all quartiles (Q1 to Q4), no notable differences were observed in the similarity of papers between high impact and lower impact journals (Menshawey et al., 2023).

A qualitative analysis comparing six features and performance metrics across five text plagiarism detection software programs reveals that iThenticate holds the second position in the overall ranking among these tools (Ali et al., 2011). An evaluation of nine text-matching software programs for plagiarism detection highlighted Turnitin and iThenticate as the top-performing tools in key functional categories, focusing on originality, image-text, citations, character-like symbols, and invisible characters in five documents (Elkhatat et al., 2021). A comparison of plagiarism detection across 100 articles using two paid and two free software programs showed that the free tools had similar and lower average rankings compared to the paid ones (Anil et al., 2023). In text similarity evaluations, paid plagiarism detection software proves to be more effective than unpaid options (Jain et al., 2016; Mahian et al., 2017).

Conclusions

In conclusion, the comparative analysis of the similarity index among iThenticate, Ouriginal, and Turnitin plagiarism detection software reveals distinct differences in their performance. Turnitin plagiarism detection software demonstrated the highest average similarity detection, followed by iThenticate plagiarism detection software, and then Ouriginal plagiarism detection software. In this study, we indicate varying abilities in detecting document similarity among the tools, emphasizing the need for careful selection of plagiarism detection software for maintaining work integrity and reliability in academic and professional environments.

Conflicts of interest - The authors declare no conflicts of interest.

Funding: None

References

- Adithan, C., Surendiran, A. (2018). Plagiarism Software. In: Parija, S., Kate, V. (eds) *Thesis Writing for Master's and Ph.D. Program*. Springer, Singapore. https://doi.org/10.1007/978-981-13-0890-1_28
- Ali, A. M. E. T., Abdulla, H. M. D., & Snasel, V. (2011). Overview and comparison of plagiarism detection tools [Poster presentation]. Annual International Workshop on Databases, TExts, Specifications and Objects (Dateso), 161–172, Pisek, Czech Republic. <https://ceur-ws.org/Vol-706/poster22.pdf>

- Andronic, O., Bolocan, A., Păduraru, D. N., Ion, D., & Musat, F. (2022). How much do Romanian medical students know about research ethics? A survey. *European Science Editing*, 48, 1-10. <https://doi.org/10.3897/ese.2022.e7626>
- Anil, A., Saravanan, A., Singh, S., Shamim, M. A., Tiwari, K., Lal, H., Seshatri, S., Gomaz, S. B., Karat, T. P., Dwivedi, P., Varthya, S. B., Kaur, R. J., Satapathy, P., Padhi, B. K., Gaidhane, S., Patil, M., Khatib, M. N., Barboza, J. J., & Sah, R. (2023). Are paid tools worth the cost? A prospective cross-over study to find the right tool for plagiarism detection. *Heliyon*, 9(9), 1-11. <https://doi.org/10.1016/j.heliyon.2023.e19194>
- Aronson J. K. (2007). Plagiarism - please don't copy. *British Journal of Clinical Pharmacology*, 64(4), 403-405. <https://doi.org/10.1111/j.1365-2125.2007.03042.x>
- Baskaran, S., Agarwal, A., Panner Selvam, M. K., Henkel, R., Durairajanayagam, D., Leisegang, K., Majzoub, A., Singh, D., & Khalafalla, K. (2019). Is there plagiarism in the most influential publications in the field of andrology?. *Andrologia*, 51(10), 1-7. <https://doi.org/10.1111/and.13405>
- Bhattacharya, A. (2006). Plagiarism: The scientific and literary theft. *Indian Journal of Dermatology*, 51(2), 87-88. <https://www.e-ijd.org/text.asp?2006/51/2/87/26925>
- Chung, S., Lee, J., Lee, Y., Park, H. Y., & Kim, D. (2017). Characteristics of the similarity index in a Korean medical journal. *Korean Journal of Anesthesiology*, 70(3), 327-334. <https://doi.org/10.4097/kjae.2017.70.3.327>
- Condurache, I.-A., & Bolboacă, S. D. (2022). Comparison of plagiarism detection performance between some commercial and free software. *Applied Medical Informatics*, 44(2), 73-86. <https://ami.info.umfcluj.ro/index.php/AMI/article/view/904>
- Debnath, J. (2016). Plagiarism: A silent epidemic in scientific writing - Reasons, recognition and remedies. *Medical Journal Armed Forces India*, 72(2), 164-167. <https://doi.org/10.1016/j.mjafi.2016.03.010>
- Elkhatat, A. M., Elsaid, K., & Almeer, S. (2021). Some students plagiarism tricks, and tips for effective check. *International Journal for Educational Integrity*, 17(15), 1-12. <https://doi.org/10.1007/s40979-021-00082-w>
- Garg, K. C., & Nagpal, R. (2023). Role of the editor in limiting plagiarism or similarity in scholarly journal manuscripts. *Journal of Scientometric Research*, 12(2), 505-508. <https://doi.org/10.5530/jscires.12.2.047>
- Green, S. P. (2002). Plagiarism, norms, and the limits of theft law: Some observations on the use of criminal sanctions in enforcing intellectual property rights. *Hastings Law Journal*, 54(1), 167-242. <http://dx.doi.org/10.2139/ssrn.315562>
- Harris, R. A. (2001). *The plagiarism handbook: Strategies for preventing, detecting, and dealing with plagiarism*. Los Angeles (Calif.): Pyczak Publishing
- Helgesson, G., & Eriksson, S. (2015). Plagiarism in research. *Medicine, Health Care and Philosophy*, 18(1), 91-101. <https://doi.org/10.1007/s11019-014-9583-8>
- iThenticate. (2023). Retrieved 11 2023, 27, from Web Page: <https://www.ithenticate.com/>
- Jain, H., Das, S., & Garg, A. (2016). Google versus other text similarity tools in detection of plagiarism: A pilot study in the journal of clinical and diagnostic research. *European Science Editing*, 42(4), 87-90. <https://doi.org/10.20316/ESE.2016.42.012>
- Joob, B., & Wiwanitkit, V. (2018). Plagiarism: Either intentional or unintentional, it is still plagiarism!. *Perspectives in Clinical Research*, 9(3), 151. https://doi.org/10.4103/picr.PICR_17_18
- Joy, M.S., & Luck, M. (1999). Plagiarism in programming assignments. *IEEE Transactions on Education*, 42, 129-133. <https://doi.org/10.1109/13.762946>
- Kadam D. (2018). Academic integrity and plagiarism: The new regulations in India. *Indian Journal of Plastic Surgery : Official Publication of the Association of Plastic Surgeons of India*, 51(2), 109-110. https://doi.org/10.4103/ijps.IJPS_208_18
- Kale, S., T. (2019). Anti-plagiarism software check to the doctoral theses in Indian universities and institutions. *Plagiarism: An International Reader*. Chapter-22, 263-273, Ess Ess Publications, New Delhi, India
- Khadilkar, S. S. (2018). The plague of plagiarism: Prevention and cure!!!. *Journal of Obstetrics and Gynaecology of India*, 68(6), 425-431. <https://doi.org/10.1007/s13224-018-1182-9>
- Kirtania, D. K. (2023). ChatGPT generated content and similarity index in chemistry & allied sciences. *Journal of Chemical Information and Modeling*, 10.1021/acs.jcim.3c01110. Advance online publication. <https://doi.org/10.1021/acs.jcim.3c01110>
- Luksanapruksa, P., & Millhouse, P. W. (2016). Guidelines on what constitutes plagiarism and electronic tools to detect it. *Clinical Spine Surgery*, 29(3), 119-120. <https://doi.org/10.1097/BSD.0000000000000371>
- Mahian, O., Treutwein, M., Estellé, P., Wongwises, S., Wen, D., Lorenzini, G., ... Sahin, A. (2017). Measurement of similarity in academic contexts. *Publications*, 5(3), 18. <https://doi.org/10.3390/publications5030018>

- Manley, S. (2023). The use of text-matching software's similarity scores. *Accountability in Research*, 30(4), 219–245. <https://doi.org/10.1080/08989621.2021.1986018>
- Martin, B. (2004). Plagiarism: Policy against Cheating or Policy for Learning?. <https://ro.uow.edu.au/artspapers/78>
- McIntosh, C. (2013). *Cambridge Advanced Learner's Dictionary* (4th ed.). Cambridge, UK, Cambridge University Press. <https://cir.nii.ac.jp/crid/1130000795527672192>
- Memon, A. R., & Mavrinac, M. (2020). Knowledge, attitudes, and practices of plagiarism as reported by participants completing the authorAID MOOC on research writing. *Science and Engineering Ethics*, 26(2), 1067–1088. <https://doi.org/10.1007/s11948-020-00198-1>
- Menshawey, R., Menshawey, E., Mitkees, A., & Mahamud, B. A. (2023). A plagiarism paperdemic: Determining plagiarism among COVID-19 articles in infectious disease journals between 2020 and 2021. *Bulletin of the National Research Centre*, 47(1), 1-11. <https://doi.org/10.1186/s42269-023-01129-3>
- Meo, S. A., & Talha, M. (2019). Turnitin: Is it a text matching or plagiarism detection tool?. *Saudi Journal of Anaesthesia*, 13(1), 48–51. https://doi.org/10.4103/sja.SJA_772_18
- Muşat, F., Păduraru, D. N., Bolocan, A., Ion, D., Constantinescu, A., & Andronic, O. (2023). Plagiarism detection in manuscripts submitted to the journal of surgical sciences between 2020 and 2021: A case study. *Science Editing*, 10(2), 149-153. <https://doi.org/10.6087/kcse.313>
- Narayan, S. S. (2023). Plagiarism vs similarity index: A critical insight. *German Journal of Pharmaceuticals and Biomaterials*, 1(4), 1–2. <https://doi.org/10.5530/gjpb.2022.4.14>
- Nkotagu, M. H. (2022). *Development of an Algorithm for Plagiarism Detection* (Masters Theses and Dissertations, NM-AIST). <https://doi.org/10.58694/20.500.12479/1624>
- Ouriginal. (2023). Retrieved 11 2023, 27, from Web Page: <https://www.ouriginal.com/>
- Patra, S. K., & Kirtania, D. K. (2023). OpenAI ChatGPT generated content and similarity index: A study of selected terms from the library & information science. *Annals of Library and Information Studies*, 70(2), 99-101. <https://doi.org/10.56042/alis.v70i2.118>
- Peh, W. C., & Arokiasamy, J. (2008). Plagiarism: A joint statement from the Singapore medical journal and the medical journal of Malaysia. *Singapore Medical Journal*, 49(12), 965–966
- Scanlon, P. M. (2003). Student online plagiarism: How do we respond? *College Teaching*, 51(4), 161–165. <http://www.jstor.org/stable/27559159>
- Scanlon, P. M., & Neumann, D. R. (2002). Internet plagiarism among college students. *Journal of College Student Development*, 43(3), 374-385
- Smart, P., & Gaston, T. (2019). How prevalent are plagiarized submissions? Global survey of editors. *Learned Publishing*, 32(1), 47-56. <https://doi.org/10.1002/leap.1218>
- Swaan P. W. (2010). Publication ethics—A guide for submitting manuscripts to Pharmaceutical Research. *Pharmaceutical Research*, 27(9), 1757–1758. <https://doi.org/10.1007/s11095-010-0188-5>
- Turnitin. (2023). Retrieved 11 2023, 27, from Web Page: <https://www.turnitin.com/>
- Usta, U., & Koçak, Z. (2021). Authors and editors of scholarly journals should be aware of plagiarism and its consequences. *Balkan Medical Journal*, 38(6), 321–323. <https://doi.org/10.5152/balkanmedj.2021.21001>