

Online Plagiarism Detection Tools in the Digital Age: A Review

Vandana Chandere¹, S Satish² and R Lakshminarayanan³

¹ Senior Technical Officer–(2), ICMR-National Institute of Virology, Pune,
Maharashtra

² Technical Officer- C, ICMR-National Institute of Epidemiology, Chennai,
Tamil Nadu

³ ADG (Admin), Indian Council of Medical Research, New Delhi
Corresponding author's email: yessatish@gmail.com

Abstract. Academic and research institutes are engaged in teaching and formulation of new research. The major problem in research publication is plagiarism, copying of published work without proper citation of source. Academic integrity breaches downloading, copying, and pasting paragraphs whole or part of assignments from the Internet. University Grant Commission (UGC) has introduced the "Prevention of Plagiarism in Higher Education Institutes" regulation in 2018 to prevent plagiarism submitted by students, staff, and researchers. UGC has determined four levels of plagiarism based on percentages, using the anti-plagiarism checker software. Plagiarism has become a significant concern, and researchers seek to protect their work. Different commercial plagiarism checker tools are available to check the originality of content, e.g., Turnitin, URKUND, iThenticate, etc. This review will discuss plagiarism, types of plagiarism, reasons for plagiarism among the research community, University Grant Commission rules and regulations for preventing plagiarism, plagiarism levels, plagiarism detection, and how the plagiarism checker software works. We have also discussed the comparison between various commercial and online free plagiarism detecting and preventing software in detail, the librarian's role for avoiding plagiarism, etc. Besides, guidelines, suggestions, the percentage for avoiding plagiarism, awareness among the research community discussed.

Keywords: Plagiarism detection tools, Misconduct, plagiarism software, UGC regulation, commercial and online free plagiarism tool, Information and Communication Technology

1. Introduction

Human seeks to quantify and measure everything that comes across. This trending tenet 'publish or perish' has put immense pressure on scientists, teachers, researchers, and even the students to publish more and more papers in their bucket^[1]. Everything is related to career, academic promotion, job, salary, respectability in literary society. It is a neurotic obsession to publish, which has led to the increasing market of predatory journals worldwide^[9].

The Internet has revolutionized the way of accessing, organizing, managing, retrieval, and disseminating information. It provides different types of tools and services for accessing electronic resources all over the world. In this age of the Internet, instant solutions, i.e., 'cut and paste' from other sources like websites, e-journals, ebooks, thesis, dissertations, reports, have become the usual practice among the academic and scientific community for writing a manuscript. Copying of published work without proper acknowledgment of the source is called plagiarism^[6]. In other words, "plagiarism is a work of deliberately publishing the content or work of someone and passing on as a work of his/her own"^[16]. Plagiarism or unreasonable copying in journal articles is a severe growing problem in the country^[18]. The word plagiarism originates from the Latin word "Plagiarism," which means kidnapper, used in the 17th century^[3]. Plagiarism is the representation of another person's work as one's own^[5]. Plagiarism is copying the content of another's work or research to save the time required in conducting research.

2. Purpose

The primary purpose is to create awareness among researchers, students, and staff about the various anti-plagiarism software, which is now mandatory for any research and publication. Indian Council of Medical Research (ICMR), New Delhi, has formulated guidelines on publication ethics. This guideline will support to improve the quality of publication in scientific/academic research.

3. Types of Plagiarism in Research

Plagiarism is an offense that should be avoided by the students, staff, and researchers. Plagiarism broadly classified into different types, which a researcher must remain fully aware of are mentioned below in the Table. 1 ^{[13][15]}.

Table 1. Types of Plagiarism in Research

	Types of Plagiarism	Description
1.	Deliberate Plagiarism	Deliberate plagiarism occurs as an intentional act of attempting to use someone else's work as one's own. It is the most used form of plagiarism and can be detected by using online software.
2.	Paraphrasing	Paraphrasing occurs when a researcher summarizes an idea taken from another source and does not cite the author's name and provide the own corresponding reference.
3.	Patchwork Paraphrasing	It is also known as mosaic plagiarism or total plagiarism. It means copying and pasting text from two or more resources to create a new text. They use to reword the sourced material while keeping the structure of the original texts by the researchers.
4.	Stitching Sources	This type of plagiarism is difficult to detect. All source lists are correct and accurately cited, but the researcher cannot critically analyze the source texts to produce their work.
5.	Global Plagiarism	It means to pay to write theses or research for using someone else's work or buying assignments to show own work.
6.	Artistic Plagiarism	Using research in different media like text, images, audio, video, etc.
7.	Verbatim Plagiarism (Copy & Paste)	This type of plagiarism is easy to detect if the text is copied directly from the online source without properly citing the information. The majority of words, structure even formatting find the same in this type of plagiarism.
8.	Self-Plagiarism	Self-Plagiarism refers to reuse their work in generating new research and do not reference it appropriately.
9.	Bluffing/Fake Plagiarism	This plagiarism after reading books, journal articles, manuals, standards, reports, creates a new idea and is shown that they are different from them, but in reality, these thoughts are the same.
10.	Citing Incorrectly	The first step in avoiding plagiarism is citing your sources. Reference appropriately cited as defined reference style like APA, MLA, Chicago Reference style, etc. If the above type does not follow the citing, it comes under plagiarism.
11.	Citing a non-existent source	It is also a form of plagiarism. It means citing papers which are unable to find or locate properly.

4. Reasons for Plagiarism among Research Community

Over the past decades, plagiarism identified in various forms mentioned above that occur in research

institutes and even in academic higher education. There are several reasons discussed below in Figure.1 [8][11]

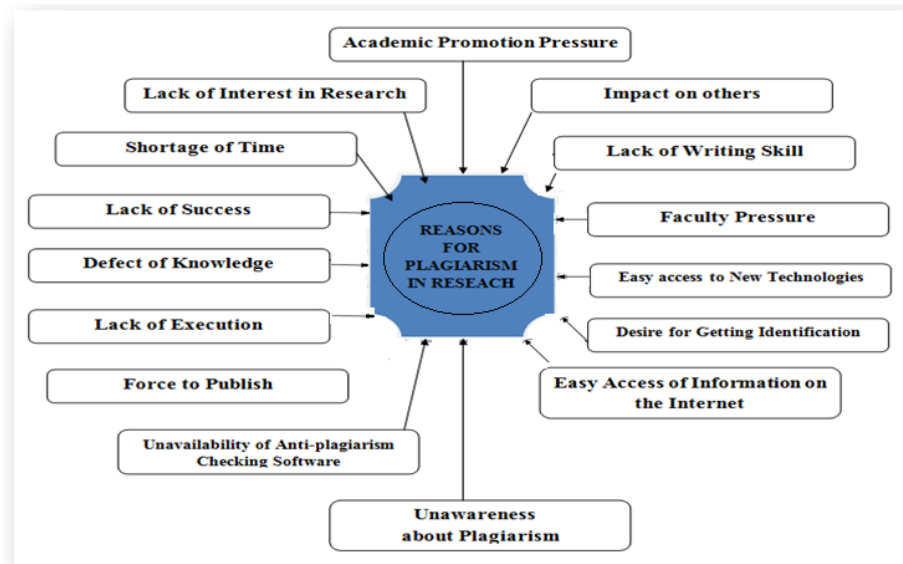


Figure 1: Reasons for Plagiarism among Research Community

5. University Grant Commission Rules & Regulation for Preventing Plagiarism in Academic Institutes in India

UGC has been set up for the coordination, determination, and set up of standards of higher education in India. It recognizes universities and disbursement of funds to all recognized colleges and universities. UGC has set some regulations, mandates to be followed by the universities, and maintained education standards. UGC has established a consortium for academic and research ethics (CARE) to keep a reference list of quality journals to avoid predatory journal publications.

UGC Regulation 2018: UGC has announced regulation for preventing plagiarism in the 530th meeting of UGC on "Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions Regulations, 2018" June 2018¹⁶. The main objectives of these regulations are :

- To create awareness regarding the conduct of research, dissertations, project work, assignments, and thesis including plagiarism among the academic and scientific community
- Set up an institutional mechanism through education and training to facilitate responsible research, promote academic integrity, and deter plagiarism.
- To set up a system and set up a mechanism to detect and prevent plagiarism

6.Levels of Plagiarism & Penalty:

Table 2. Levels of Plagiarism & Penalty as defined by UGC

S.N.	Levels	Description	Penalties
1.	Level 0	10% similarity or less	No penalty

2.	Level 1	10% - 40 % similarity	Submit a revised script at a certain time (Not more than 6 Months)
3.	Level 2	40% - 60 % similarity	Submit a revised script (Not more than 16 Months)
4.	Level 3	More than 60% similarity	Cancellation of Registration

Source: UGC Notification Published in The Gazette of India on 23 July 2018

Plagiarism percentage in various subjects may vary in research. A portion of similarity content will be high in descriptive study, narrating the developmental aspects, tracing the concept, and historical background. The projecting view of authors, researchers, and research on some subjects like philosophy, literature, religion, and biography is likely to have more plagiarism due to such research relying much on existing literature and old aspects.

7. Departmental Academic Integrity Panel (DAIP)

Higher Education Institutes shall compose a Departmental Academic Integrity Panel (DAIP) for the follow-up. The regulation on the institute level is mentioned below:

- Chairman - Director, Principal or Head of the Department
- Member - Teachers, Scientists, and others from outside the department.
- Member - A person well efficient with the anti-plagiarism tool.

8. Detection Methods of Plagiarism

Plagiarism detection is done manually or by using computer software. Manual detection takes more time and requires more effort. It is useful when one or two pages are available to compare, but many documents are impossible through this method. Computer software-based detection allows many online records collections like webpages, ebooks, e-journals, databases, etc. It is easy to compare using particular software^[10]. There are many methods to detect plagiarism mentioned in Figure 2 and the workflow of online plagiarism software are described below in Figure. 3.

Figure 2. Plagiarism Detection Methods and Techniques

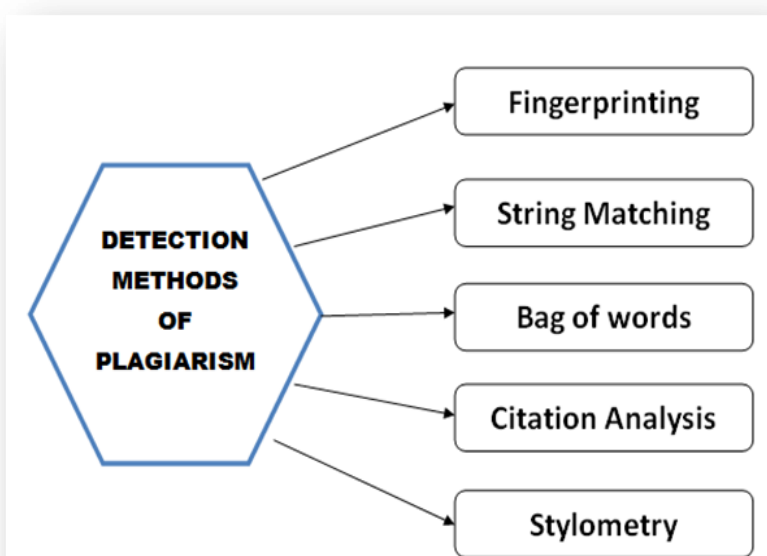
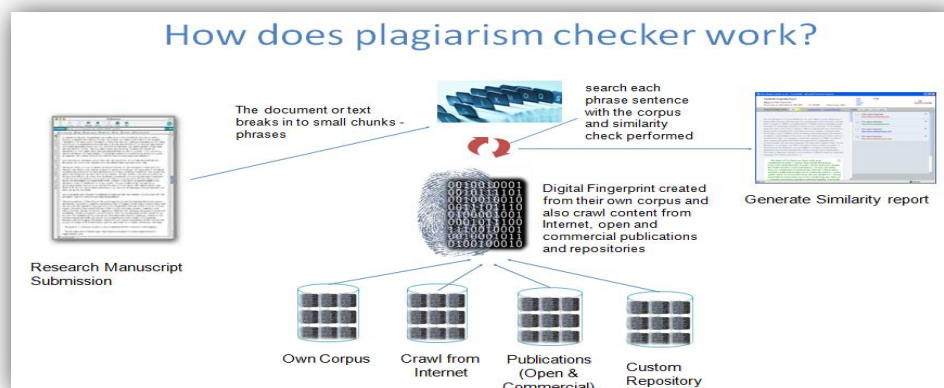


Figure 3. How does plagiarism checker online software work



Source: "Plagiarism Detection: Products, Services, and Technologies" <https://slideplayer.com/slide/14049829>"

9. Plagiarism Detection and Prevention software

Anti-plagiarism software was used in 1993 at Harvard USA to check students' projects. There are currently many software available: subscription-based or partial, online free or freemium basis plagiarism checker. Some online free software checks plagiarism-free, but their word limit varies. They do not provide an accurate report as required by the researchers. Subscription-based software like Turnitin, Urkund, IThenticate gives better results and generates a certificate of a percentage of plagiarism for thesis and manuscripts. Some software support multiple file formats and various languages like Copyleaks, Viper, etc., while others have additional features like proofreader & grammar check, vocabulary builder, etc. Lists of anti-plagiarism software with details are enlisted in Table 3^[4]:

Table 3. List of Online Anti-plagiarism Softwares

S.N.	Name of the Software	Mode of Access	Year	Developer	Website
1.	Turnitin	Subscription	1997	iParadigms, USA	https://www.turnitin.com/
2.	iThenticate	Subscription	2004	iParadigms, USA	http://www.ithenticate.com/
3.	Urkund	Subscription	2000	PrioInfocenter/ Urkund, Sweedan	https://www.urkund.com/
4.	PlagScan	Subscription	2008	PlagScan GmbH, Germany	https://www.plagscan.com/en/

5.	Viper	Free	-	ScanMyEssay, Viper, England	https://www.scanmyessay.com/
6.	Grammarly	Freemium	2009	Grammarly, Inc.	https://www.grammarly.com/plagiarism-checker
7.	Copyleaks	Subscription	-	COPYLEAKS	https://copyleaks.com/
8.	Unicheck	Subscription	-	Unicheck	https://unicheck.com/
9.	PlagTracker	Freemium	2011	Devellar, Ukrainian	https://www.plagtracker.com/
10.	Duplichecker	Free	-	DupliChecker	https://www.duplichecker.com/
11.	Plagiarism Checker	Free	-	Smallseotools	https://smallseotools.com/plagiarism-checker/
12.	Plagium	Freemium	-	Septet Systems	https://www.plagium.com/en/plagiarismchecker
13.	Copyscape	Freemium	2004	Indigo Stream Technologies, Ltd.	https://www.copyscape.com/

Source: https://en.wikipedia.org/wiki/Comparison_of_anti-plagiarism_software

Some software has a unique feature that differentiates them from other software like Copyscape targeted at website managers and helps find copied pages on the web.

10. Comparison among Anti-plagiarism Software based on its Characteristics:

SN.	Particulars	Turnitin	URKUND	Viper	PlagScan	Ithenticate
-----	-------------	----------	--------	-------	----------	-------------

1.	Scope of search	Web pages, scholarly journals, other contents, student paper database like ePathshala, shodhganga, vidhyanidhi, other databases like Pubmed, ERMED, JGate, Indmed, Scopus, ProQuest, Newspapers, etc.	Online Books, Journals, Websites, News, Internet, Student database	10 Billion Sources of Books and Journals	Web documents, journals, internal archives (The Internet' with 14 billions of digital content)	Web page, Scholarly content items, Journal Databases, books, Standards, Blogs, Online News, conference proceedings, etc.
2.	Supporting of Other Languages	19 languages.	Yes	Supports 56 languages	Yes	Yes
3.	Multiple File Format Supporting	MS Word, Word XML, PDF, HTML, PPT, RTF	Pdf, MS Word, PPTs, HTML, txt etc	Microsoft Word, PDF, Open Office, and Google Docs format	Yes	PDF, MS Word, PPTs, HTML etc.
4.	Analysis of time	Average	Fast	The average premium scan time of fewer than 30 seconds	Average	Average
5.	Bulk Upload	Yes	-	One at a time	-	-
6.	Direct Source Matching and Provides the URL of the Source	Yes	Yes	Yes	Yes	Yes

7.	Interpret Plagiarism Score	Yes	Yes	Yes	Yes	Yes
8.	Downloadable Plagiarism Report	Yes	Yes	Yes	Yes	Yes
9.	Check against own work	Yes	-	Yes	Yes	-
10.	Citation Verification	Yes	-	-	-	-
11.	Limitation of Pages or data	400 Pages or 40 MB Maximum	-	-	-	400 pages 40 MB (25000 Pages)
12.	Available for Various type of Users	Yes	Yes	Yes	Yes	Yes
13.	Available for Single User	Yes (Write Check)	No	Yes	Yes	Yes
14.	Support	Yes	Yes	Yes	Yes	Yes
15.	API & Plug-In	No	Yes	-	Yes	Yes

Table 4: Comparison between Various Online Plagiarism Checker Softwares

11. Educate to prevent plagiarism:

In the fight against plagiarism in an academic/ research Institution, educate and create awareness among students, teachers, scientists, and educate them about the types of plagiarism and to identify the plagiarism. Moreover, they must use a three-step process to fight against plagiarism, which involves taking steps before, during, and after creating the research^[14]. Use the online plagiarism checker software before submitting the thesis, dissertation, and manuscripts and carefully cite the source.

12. Role of the Librarian to Prevent Plagiarism

Nowadays, most institutes' librarians work as members of the Departmental Academic Integrity Panel (DAIP) to check plagiarism at the institute level. UGC has specified three members in DAIP which; one member should be familiar with plagiarism checker software, and the librarian can be part of this panel. The librarian may have additional roles in the context of prevention of plagiarism, as mentioned below:

- Conducting orientation program for the researchers, students, scientists, staff, teachers, etc.
- Organize seminars/ workshops and lectures from time to time for the users and Library professions.
- Purchase of anti-plagiarism software to the users at the institute level.

13. Guidelines/ suggestions for avoiding plagiarism

- Use the reference of copied, converted, or paraphrased material and use quotation marks.
- Give proper acknowledgments.
- Obtain permission from the publisher and author for the quotations.
- Use the standard reference style like MLA, APA, Chicago, etc., for citing the source.

14. Conclusions

In academic institutes, plagiarism detection and prevention has become one of the significant challenges^[2]. The case of plagiarism in research has become usual among researchers. UGC has prepared regulations for researchers and students in academic/research institutes to control plagiarism and follow UGC's research ethics and guidelines to ensure proper implementation and compliance of these regulations by the respective institutes. This paper has compared various online anti-plagiarism software, plagiarism detection software for research and academic institutes, and the community in detail. This paper may help researchers, students, and staff submit their publication without plagiarism and help a librarian select and procurement anti-plagiarism software accordingly.

References

- [1] Ahmed RKA 2015 Overview of Different Plagiarism Detection Tools. International Journal of Futuristic Trends in Engineering and Technology, **2** p1-3.
- [2] Bretag T 2013 Challenges in addressing plagiarism in education. PLoS Med, 10. e1001574. Available from : <http://doi:10.1371/journal.pmed.1001574>.
- [3] Chauhan SK 2018 Research on Plagiarism in India during 2002-2016: a bibliometric analysis. DESIDOC J Lib Inf Tech, **38** (2) 69-74. Available from: <http://doi:10.14429/djlit.38.2.12298>.
- [4] Comparison of anti-plagiarism software. Available from: https://en.wikipedia.org/wiki/Comparison_of_anti-plagiarism_software. Accessed 21-10-2020
- [5] Gupta R. Top 10 Free Online Plagiarism Checkers Tools in 2017. Available from: <https://yourstory.com/mystory/f65e68ea8c-top-10-free-online-plagiarism-checkers-tools-in-2017>. Accessed on 25 September 2020.
- [6] Kharat R, Chavan PM, Jadhav V, Rakibe K 2013 Semantically detecting plagiarism for research papers. International Journal of Engineering Research and Applications, **3**(3) pp 77-80.
- [7] Manohar SM, Vajjaha M 2014. Plagiarism Detection in Computer Science. International journal of computer science trends and technology, **2**(5) pp 171-4.
- [8] Mishra R, Gautam VK 2017. Issue, and challenges of plagiarism in the digital environment: a conceptual analysis. Journal of Advancement of Library Science. **4**(3) pp 65-9.
- [9] Misra DP, Ravindran V, Wakhlu A, Sharma A, Agarwal V, Negi VS 2017. Plagiarism: a Viewpoint from India. *J Korean Med Sci*, **32** pp 1734-5. Available from <http://doi.org/10.3346/jkms.2017.32.11.1734>.
- [10] Paul C 2003. Old, and new challenges in automatic plagiarism detection. In: National

- Plagiarism Advisory Service. pp 391-407.
- [11] Plagiarism and copyright Reasons Students Plagiarize or Cheat. Available from <https://www.rit.edu/twc/academicintegrity/reasons-students-plagiarize-or-cheat>. Accessed 17-10-2020)
- [12] Rodriguez V. Plagiarism, its consequences, and how to avoid it. Available from <https://pdfs.semanticscholar.org/b70b/a0de83d78d17171cb5b427b3f377756a1f8e.pdf>. Accessed 19-11-2020
- [13] Streefkerk R. Types of plagiarism. Available from <https://www.scribbr.com/plagiarism/types-of-plagiarism/>. Accessed 12-11-2020
- [14] Thomas, R. How does India move on fighting plagiarism. Available from <https://yourstory.com/mystory/how-does-india-move-on-fighting-plagiarism-s4unsj1267>. Accessed 20-11-2020
- [15] Types of Plagiarism. Available from <https://www.scanmyessay.com/plagiarism/types-of-plagiarism.php>. Accessed 12-11-2020
- [16] University Grants Commission 2017 (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Education Institutions) Regulations. Available from https://www.ugc.ac.in/pdfnews/8864815_UGC-Public-Notice-on-Draft-UGC-Regulations Accessed 16-10-2020)
- [17] Weber-Wulff R. Plagiarism detectors are a crutch and a problem. Available from <https://www.nature.com/articles/d41586-019-00893-5>. Accessed 15-11-2020
- [18] Yuehong Z, Xiaoyan JIA (2012). A survey on the use of cross-checking for detecting plagiarism in journal articles. *Learned Publishing*, **25** pp 292-307. Available from <http://doi:10.1087/20120408>.