Short Communication

Challenges in using anti-plagiarism software: a case study

Shipra Awasthi[a] and Manorama Tripathi[b]

[a]Assistant Librarian, Jawaharlal Nehru University, New Delhi,
E-mail: shipra2010@gmail.com
[b]Librarian, Jawaharlal Nehru University, New Delhi,
E-mail: manoramatripathi2@yahoo.com

The paper highlights the coverage, features, functionalities and limitations of different anti-plagiarism software used at the Jawaharlal Nehru University, New Delhi. While anti-plagiarism software can enhance the quality of writing, the study reiterates that anti-plagiarism software should be used in conjunction with human intelligence.

Introduction

Higher educational institutions worldwide face the challenge of maintaining and ensuring honesty in education and research as often plagiarism, contract cheating, copying and submitting past works and other forms of misconduct tend to take place.[1] The UGC regulations 2018 have mandated the use of anti-plagiarism software to scrutinise research and ensure its originality.[2] There are various proprietary and open source software available for checking similarity in the contents. Dr B R Ambedkar Central Library, Jawaharlal Nehru University, uses Turnitin, Urkund and Drillbit for scanning documents for originality. It also used Check-For-Plag on a free trial basis for one year. The features of the four software tools are given in Annexure I.

The university library regularly organises orientation programmes to sensitise users on misconduct issues in education and research. It holds sessions on how to publish in peer-reviewed journals, identify and avoid predatory journals and conferences, use reference management tools, anti-plagiarism software for research and so forth.

The present study compares the features, functionalities, coverage, levels of access, and the performance of four anti-plagiarism software. The paper dwells on the different challenges which are encountered in using these anti-plagiarism software.

The JNU Library checks theses, dissertations, research manuscripts through Turnitin, while other documents like term papers, assignments, etc are checked through Ouriginal (Urkund) to ensure the maximum utilization of all the software. The four antiplagiarism software were examined on the parameters of accepted file formats, file size, coverage of the database, the file format of the originality report, repository provision, and grading and grammar checking features. The same documents were not checked through all the four antiplagiarism software. The authors intend to undertake the screening of the same documents through the four antiplagiarism in future research.

The anti-plagiarism software have limitations which have been highlighted in the succeeding paragraphs.

TURNITIN

The report is downloaded in PDF and HTML formats, and the downloading process in PDF is slow. The report gives false positive for affiliation, title, and authors name despite setting the exclusion of 14 consecutive words. There is no search interface to look for any file by title or author's name. The listing of files or assignments does not follow any chronological order. When the report is generated, it only lists the matched sources. When the scholars exclude any source, it does not appear in the report. We found difficulties in comprehending the report. The variation in the overall similarity index and the similarity index with sources lacks clarity. For instance the similarity index with sources (Internet sources, publications, and student papers) is 0%, but the overall similarity index is 21%.

Further bifurcation of sources displayed <1%. If we add up all the sources, it does not total to 21%. The originality report excludes 14 words but not 14 consecutive words as per UGC regulations. The toggling of the filter option changes the similarity index of the content, as displayed in Figs. 1 and 2. The similarity by sources like the internet, students' papers and publications are a relative index and does not portray a clear picture. The same content uploaded in two different Turnitin accounts generated different reports.

Turnitin provides instructors (a feature) the ability to exclude small matches, common knowledge (universal rules, formulae, methods and materials and contents) which may be similar but not plagiarised. When the sources are excluded, the system does not consider and count them while generating an originality report. Ideally, the originality report should specify the excluded sources.

E-rater is the tool to facilitate the researchers with grammar checking. This module assists the research community in improving the quality of the content with the incorporation of the correct language. The e-rater feature can't run on a document that is more than 64,000 characters.

Ouriginal (URKUND)

URKUND has changed its name to Ouriginal in March 2021. The faculty may have their account (called receiver accounts), and students (need to be registered as submitters) can directly submit to the concerned faculty member. It supports regional languages but in UNICODE only. The report is generated in PDF format. Each source in the report is listed in the matched sources list. The list is divided into two- Primary Sources and Alternative Sources. There is a provision to change an alternative source into a primary source. An alternative source cannot be excluded. Only primary source can be excluded. To exclude an alternative source, it should be listed as a primary source first, then it could be excluded from the list (for generating similarity).

There is variation in the similarity index shown on the application browser and shown in the downloaded report. It rounds off the similarity index provided in the report as displayed in Fig. 3 and Fig. 4.

The analysis address is required to upload the documents and can be sent through email to an analysis address. Reports are downloaded in PDF
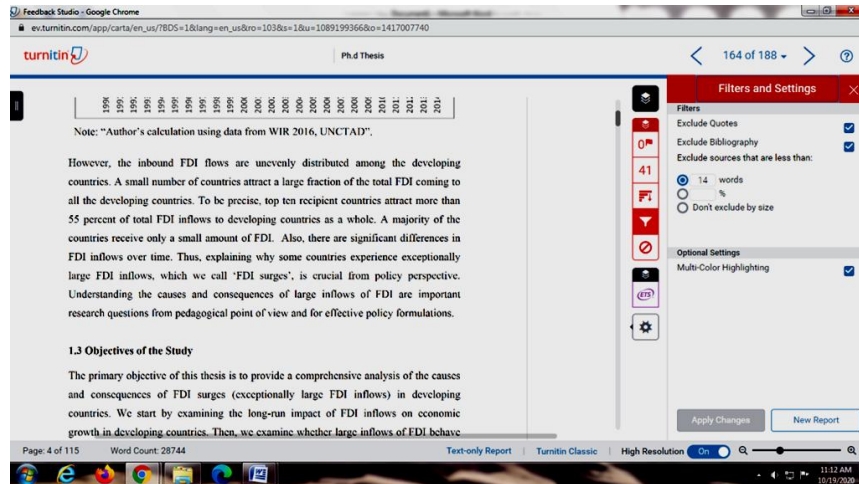

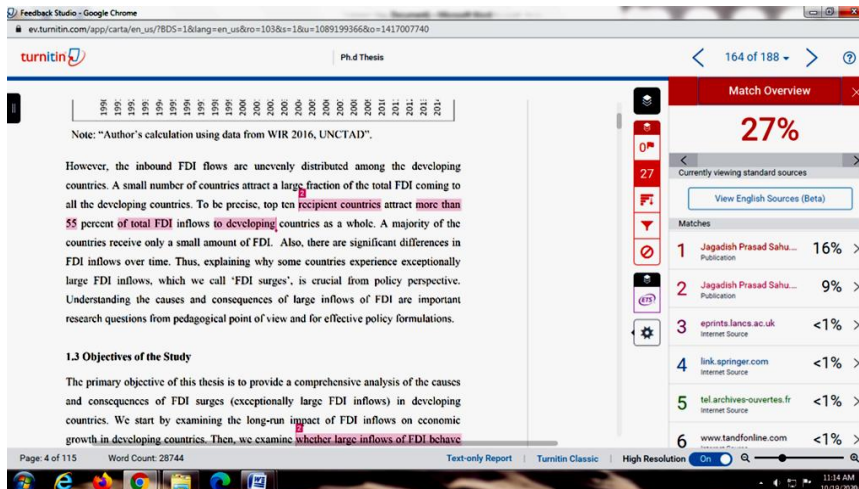
Fig. 1 — Similarity Index is 41%

Fig. 2—Similarity Index is 27%
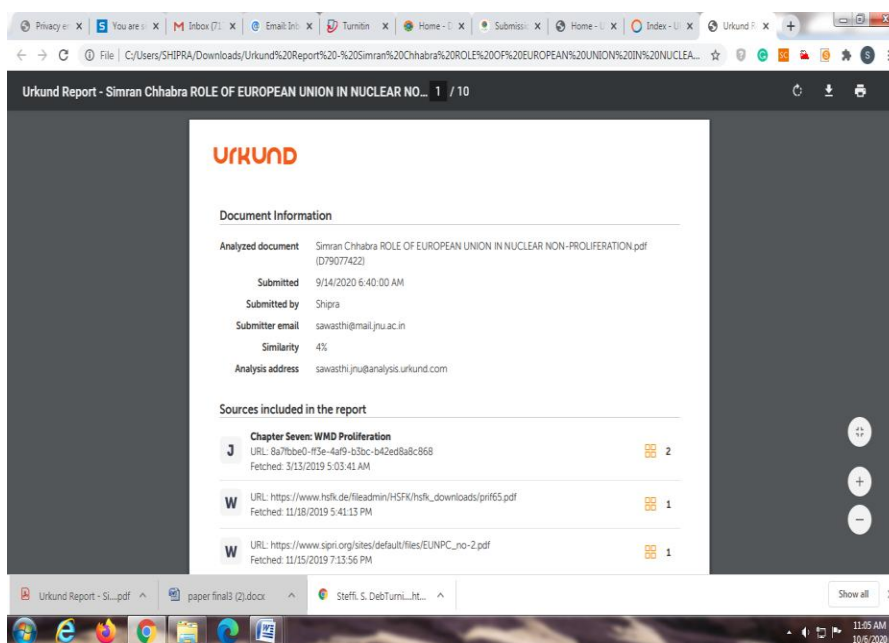
Fig. 3 — Similarity Index generated is 5%

Fig. 4 — Similarity Index generated in the report is 4%

format only, which is time-consuming. The same analysis address must be used to check the similarity of a document; if a different analysis address is used, then the similarity would be 100%. The software will match the latest version version of the content with the earlier uploaded version. Different submitter and receiver accounts should be used to check the similarity of the document. If the same account is used, then the document gets deposited in the repository. The software does not process document with formulas. It returns the message, "An unexpected error occurred when processing this document".

DRILLBIT

In the case of Indian languages, only docx file format is acceptable. The 'doc' format needs to be converted into 'docx' format before uploading. Creating a folder is mandatory; a file cannot be uploaded independently. Multiple files of regional documents cannot be uploaded.

Check-for-Plag

The JNU Library used this software on a trial basis, and encountered a few challenges. The preferences need to be set before uploading a document. There is no search box on the report page. Reports are downloaded in PDF format only, which is time-consuming.

The JNU Library checked 967, 35 and 41 documents with Turnitin, Drillbit and Ouriginal (Urkund) respectively from January 2020 to November 2020. The Library checks theses and research articles, book reviews, project proposals, etc., through Turnitin software. Other documents such as term papers, assignments etc., are checked through Ouriginal (Urkund) software. The Indian language content is checked through the Drillbit software.

The UGC regulations 2018 have defined different levels of similarity for the submission of PhD theses. Table 1 reveals the classification of documents as per the different levels mentioned in UGC regulations. The Library checked 256 and 604 theses which showed a similarity of 0- 10%, level 0 and 10 to 40%, level 1 respectively. It is generally seen that when the students learnt about the content similarity in the first round of checking, they revise and address the

Table 1 — Number of documents checked as per the levels of similarity

Different Levels	Similarity (%)	Turnitin	Ouriginal (Urkund)	Drillbit
Level 0	0-10	256	26	32
Level 1	Above 10% to 40%	604	2	7
Level 2	Above 40% to 60 %	51	0	0
Level 3	Above 60 %	48	3	0
No report generated due to error		8	4	2

problems, and in the second round of checking, the similarity goes down.

The documents categorised under Level 3 also include those that get deposited in the repository thus leading to a rise in percentage. The majority of the documents screened with Urkund and Drillbit had less than 10% of the Similarity Index (Table 1).

Turnitin software has a robust system that also identifies the hidden characters in the document. The database coverage of the other two software is not as comprehensive as that of Turnitin. The authors observed that Turnitin has more user-friendly features as compared to the other three anti-plagiarism tools.

Conclusion

Anti-plagiarism software are automated programmes and that help in improving the quality of submissions. But these tools must be used in conjunction with human intelligence and scrutiny. They are incapable of detecting certain forms of misconduct like contract cheating, manipulation of images, falsification, or data fabrication. It is strongly recommended that with the deployment of anti-

plagiarism software, sensitisation sessions be conducted to spread awareness about the detrimental effect of engaging in any kind of misconduct. The researchers need to be aware of the importance of the basic values of integrity and rigour in education and research.

References
1 Richardson M and Healy M, Examining the ethical environment in higher education, *British Educational Research Journal*, 45(6) (2019) 1089-1104.
2 University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions), (2018). Available at https://www.ugc.ac.in/pdfnews/7771545_academic-integrity-Regulation2018.pdf (Accessed on May 2021).
3 TURNITIN. Available at www.turnitin.com (Accessed on May 2021)
4 Drillbit. Available at www.drillbitplagairism.com (Accessed on May 2021)
5 Ouriginal (URKUND). Available at www.ouriginal.com (Accessed on 05 May 2021)
6 Check-for-plag. Available at www.checkforplag.com (Accessed on 05 May 2021)
7 Grammarly. Available at www.grammarly.com (Accessed on 05 May 2021)

Annexure I Anti-plagiarism software

Features	Turnitin	Ouriginal (Urkund)	Drillbit	Check- For-Plag
Vendor	TurnitIndia Education Pvt. Ltd www.turnitin.com	eGalactic www.urkund.com/	Drillbit Infotech www.drillbitplagiarism.com	Infokart India www.checkforplag.com
Max Limit	100 MB	20 MB	60 MB	Not specified on the website
Languages supported	30	30	26	17
Database coverage	70+ billion web pages, 69 million subscription articles, 17000 publishers, 1 billion student papers and 15 million pages added daily to different platforms	Internet, published documents such as journals, books, etc. and previously submitted student content	67+ Billions of Web pages, open access and commercial publishers/journals content, students repositories, institutional repositories, other open access repositories, open access articles, aggregators content, eBook and conference proceedings	Thousands of web portals, a large number of articles, web pages, and other information available on the open-access domain
Users	Fifteen thousand institutions across the globe. It provides facility to 30,000 million instructors and students	1500+ universities and institutes in India and 10,000+ globally across 100+ countries.	Not specified on the website	Not specified on the website

File Format	Microsoft Word (.doc, .docx), OpenOffice Text (.odt), WordPerfect (.wpd), PostScript (.ps/.eps), HTML, Hangul Word Processor file (.hwp), Rich Text format (.rtf), Plain text (.txt), Google Docs via Google Drive, Adobe PDF, Microsoft Powerpoint (.pptx, .ppt, .ppsx and .pps), Microsft Excel (.xls and .xlsx)	.docx, .sxw, .ppt, .pptx, .PDF, .txt, .rtf, .html, .htm,.wps, .odt	PDF,foc,docx,zip (English) Docx (Indic Languages)	doc, docx, txt and PDF
Report Format	PDF, html	PDF	PDF	PDF
E-mail notification	No	Yes	No	No
Multiple downloading of reports	No	No	Yes	No
Multiple uploads	Yes	Yes	Yes	No
QR Code	No	No	Yes	No
Grading Facility	Yes	No	Yes	No
Grammar check tool	Yes	No	Provision to integrate	No
Repository	Documents are automatically deposited in the repository if parameters are not set properly	Documents are automatically deposited in the repository	Documents can be kept out of the repository	Documents can be kept out of the repository
Levels of access	03 (Administrator, Instructor, User)	02 (Administrator & User)	02 (Administrator & User)	02 (Administrator & User)
Availability of search feature	No	Yes	Yes	Yes
Report displays excluded sources list	No Yes	Yes (Bin) No	Yes No	No No
Flag feature	Yes	No	No	No