Abstract

The paper depicts the trends in using anti-plagiarism software packages to check plagiarism in the journal articles and doctoral theses. Also explained are the aspects pertaining to the measures to be adopted for ensuring originality of research output. Experience of scanning several research reports for checking plagiarized content using anti-plagiarism software packages Urkund, Turnitin, iThenticate and Viper is explained. Authors have also dealt with the means and methods of overcoming plagiarism, especially in case of journal articles and theses submitted to the University.

Keywords: Anti-plagiarism Software; Plagiarism- Overview; Turnitin; iThenticate; Urkund.

Introduction

Since the recent past, several measures have been taken to improve the quality and originality of research contributions. In this direction, it is expected to publish papers in journals carrying International Standard Serial Number, journals carrying high impact factor and minimum percentage of similar content when scanned using anti-plagiarism software packages.

Plagiarism is nothing but borrowing the ideas of others and presenting the same as if our own. Further, when we borrow the ideas and use them as our contribution without giving credit to the authors is plagiarism. In other words, stealing someone’s ideas and later on telling lies; really an act of literary theft. Plagiarism is an act of intellectual dishonesty.

It is accepted that the attempt to avoid plagiarism is an important issue in preserving scholarly and scientific integrity. When a researcher summarizes or borrows or paraphrases or copies ideas belonging to others, he is expected to put the ideas/thoughts within quotation marks and render complete bibliographic details with names of author/s. Even changing
keywords and retaining the ideas without giving credit to authors is also treated as plagiarism. Dr. Benson Honig\textsuperscript{1} opines that significant plagiarism as "at the minimum one full paragraph, uncited and unquoted, with no reference to where it came from."

During the year 2002, there were 1.09 million research papers published in journals covering all disciplines. The number increased to 1.94 million in the year 2010. Now, there are 7 million researchers focusing on 32000 journals for publishing their papers. However, since the recent past the trend of committing research misconduct is increasing to an alarming extent. It is said\textsuperscript{2} that one out of three scientists/researchers admits to questionable research practices. The main reason behind is fabrication or falsification of data, plagiarism and mistakes in making observation and drawing inferences. Added to this, self-plagiarism is also noticed in number of cases. Since the tolerance limit of similar content is going beyond expectation, one of the popular publishers, Taylor and Francis of late, is rejecting 23\% of the papers what they receive for publication in journals. iThenticate, one of the popular commercial anti-plagiarism software is being used to scan annually 2.3 million research papers and it is highlighting to note that more than 10 million matches identified from the previously published content over the past 18 months. On account of this trend and with the intention of improving the original contribution in research output, 23 publishers insisted on plagiarism check way back in 2008 and in 2011 there were 282 publishers\textsuperscript{2} insisting on plagiarism check before accepting the paper for publication.

Elizabeth Wager \textsuperscript{5} says, "Promotion, appointments, and academic careers are relying on publication and in a way good for the publishers to open up some opportunities, I think if the pressure is too high, it will create an atmosphere in which the temptation to commit research misconduct is increased." Investigating the involvement of scholars in research misconduct is being viewed seriously and only in the USA, during 2010, 110 million dollars was spent on cost of investigation into research misconduct. Dr. Benson Honig\textsuperscript{1}, Professor at McMaster University and editor of a popular journal is of the opinion: "there is a lot of incentive for people to cut corners, try and get something published. Their job, their career and their reputation could be on the line." Often it so happens that of the 450 articles submitted for publication in a reputed journal, approximately 150 of them are being published. So the remaining 300 of them go into the great wash, and they are seen ending up in online journals--some of lesser quality. The increasing rejection rate has led to the establishment of new online journals--some of late, they are too many in number.

Plagiarism is categorized into six types and they are mentioned here below.

1. Blatant Plagiarism: Just copying the whole text without acknowledgement.
2. Pot Luck paper: Ideas borrowed from many sources
3. Word Switch: Alter keywords and phrases
4. Mosaic: paraphrase most papers
5. Self-plagiarism: Borrow from his own published papers
6. Resourceful Citer: Full of quotations and ideas of others with acknowledgement

Hence, it is essential to cite the source whenever one quotes thoughts/sentences or uses ideas of others or paraphrases works of others. When the sources are cited, there should be uniformity and consistency in rendering references, depicting original author, title, edition, place of publication, publisher, year of publication and pagination. While rendering references or bibliography, a standard referencing style has to be followed. Citations or references are useful mainly

1. To find more information on the topic
2. To verify the quotations and contents
3. To show the amount of research work
4. To support the argument of the researcher
5. To strengthen the research work done

Rendering citations will clarify who said what and where. It is our responsibility to give credit to the right author. An author is one who is responsible for thought content of the document. It is advantageous if other sources support your ideas; Citation makes your ideas stronger and more. A good paraphrasing makes the ideas of the original source fit smoothly into your paper, emphasizing the most relevant points and leaving out unrelated information. Authors often republishes a work in its entirety or portions of his previously written text while authoring a new work. Authors many a time maintain that because they are the authors, they can use the work again as they wish; they can't because self-plagiarism can infringe upon a publisher's copyright. The American Psychological Association express that plagiarism differs from self-plagiarism: Whereas plagiarism refers to the practice of claiming credit for the words, ideas, and concepts of others. Self-plagiarism refers to the practice of presenting one's own previously published work as though it were new.

Reasons of Plagiarism

During the past, interpretations and ideas originated from the brain. However, of late, it comes from Google! There are many reasons for plagiarism.

1. Lack of knowledge and poor research skills.
2. Poor expression and command over the language.
3. Pressure to publish for career advancement.
4. Desire to influence others and get recognition soon.
5. Laziness and lack of time for understanding the.

Complaints against plagiarism and measures to avoid plagiarism was there even earlier days. Attempts were being made to verify manually. In fact, it was a tedious task, requiring enormous reading and memory. Now-a-days, in the digital era, the anti-plagiarism software packages scans the web information resources, online databases, electronic books, journals, theses, report literature and proceedings of conferences and seminars, including the institutional repositories of various academic and research organizations before the percentage of similar contents is depicted. Within a time lag of 20 Minutes one can see the matches of similar content taken from various already published sources. The software
therefore, is a boon for researchers to avoid unnecessary similar content in their papers or research reports. It is also treated as an opportunity to rectify the quotations, in case of any mistakes or incompleteness. Here, the researchers will be able to trace the original author and source. The researcher will also get the complete bibliographic details of the sources from where the quotation or the ideas originated. In the absence of sources in the library, the software shows the bibliographic details of authors, title, journal, volume and issue number and pagination.

**Plagiarism Software**

Anti-plagiarism software was first used at the Harvard, USA in the year 1993 to check the plagiarized content in the student submissions of project works and assignments. Now, a large segment of academic institutions in the United States are relying upon the tool to avoid plagiarized content in research papers and reports. Initially, the paper or manuscript will be scanned. The World Wide Web and archives/databases will also be scanned to detect the similarity in content, word by word. Then it is likely you will see some red in your results. Common keywords and phrases may trigger in red. If there are complete sentences that are not original, directly taken from the already sources, then the software will identify the original source of plagiarized content that was copied. In the further steps, when you click on any red line, you can see original source for your reference and record; the tool makes provision for side by side comparison. Finally, you will be given an exact percentage that tells you how original your research paper/thesis is. Use of anti-plagiarism software is not to be mistaken for fault finding mission. In fact, it is a fact finding exercise, aiming to identify same and similar content in already published sources and thereby avoiding it, with the intention of improving the originality of research contribution.

While up-loading files of manuscripts for plagiarism check service, the document has to be on a single word or PDF file. The software accepts Word, HTML, PDF, XML, Corel Word Perfect, Rich Text Format, Adobe PostScript, Plain Text -TXT files for scanning. It is important here to highlight that the software does not scan graphs, images, photographs. Watermarks, light coloring leads to exclusion. Similar fonts used for sub-heading will not be recognized by the software.

**Popular Plagiarism Detection Software packages**

1. iThenticate, California, USA.
2. Turnitin, California, USA
3. WriteCheck, California, USA
4. Viper, England
The anti-plagiarism software scans a massive database of scholarly text, as well as billions of web content, in order to highlight instances of duplicate content and probable damage that may occur on account of copied data and research findings.

iThenticate supports 30 languages of the world. However, most of them are European languages. Arabic, Farsi, Turkish, and Korean are also included here. The anti-plagiarism from Kottayam: Nitya D'Arch scans Indian languages. Large majority of the customers for this commercial software is the researchers who have drafted research papers and theses in Malayalam.

Viper is an open source software from England. The coverage of the software is about 10 billion pages. Whereas, iThenticate covers over 40 billion² pages of information before it throws away similar content from where ideas are borrowed.

**Exclude Options in Software**

Exclude options are provided by the software. While checking for plagiarism, the entire theses has to be in single word or PDF file eliminating preliminary pages like declaration, certificate, acknowledgement, table of contents, list of charts, graphs and tables. Further, portion of appendices, comprising of glossary, questionnaires, indexes have to be omitted from the thesis while up-loading for scanning. While projecting percentage of similar content, it is essential to arrive at the percentage after excluding certain portions of the theses. Finally, the similarity report is generated after excluding the following content as well from the doctoral theses.

1. Software facilitates exclusion of text rendered within quotes.
2. When the text or paragraphs carry references in the theses, the cited text gets excluded.
3. Make provision for exclusion of phrases and key-words used in the title of the theses. These phrases often appear in the text of the theses here and there frequently. Hence, this has to be exercised in the folder option given for exclusion.
4. There is common option to exclude abstract, methods and materials (methodology) and bibliography.
5. Provision is made for exclusion of small matches comprising of up to 10 (standard) or more words that appear in the text of the theses.

6. In the software, provision is made to exclude small sources comprising of up to 10 or more words.

7. A standard format has to be followed for rendering references/bibliography. Only in such cases, the paragraphs taken from various sources get excluded.

8. Uniformity and consistency should be there while rendering bibliographic references. The references and bibliography should not be copy and pasted because neither the references and bibliography get excluded nor the cited text get excluded. Therefore, neither the text or the references be copy and pasted from Google.

The software scans quite a number of sources covering more than 40 billion pages of information before depicting the degree of similar content. Information sources made available on the World Wide Web, electronic books and journals, doctoral theses, proceedings of the seminars, conferences and workshops, institutional repositories of various academic and research organizations and their digital libraries hosted on the web.

1. Internet resources
2. Cross-Check: Proceedings of Conferences and Seminars, Research Reports
3. Publications: Books, Journals and Periodicals
4. Institutional Repositories

Of late, UGC-INFLIBNET Centre is recommending Urkund anti-plagiarism software package. In case of Urkund anti-plagiarism software, the coverage of the contents is comparatively low and more of Swedish documents. The students' submissions are included so as to check and depict the similar contents. However, the student submissions will be in Swedish language. Some of the Indian universities are scanning doctoral theses with Urkund to get report of published contents. Visvewaraya Technological University has recommended Turnitin software for the engineering colleges who offer Masters Degree courses. Here, the dissertations have to be scanned using Turnitin software and the extent of similar contents has to be within 25% - Tolerance Limit.

**Viewing Reports of Similar Content.**

In the first stage, the overall percentage of similar content will be projected. Further, down below, the percentage of similar content from Internet sources, cross-check database of journals and books will be depicted. Then the percentage of similar content taken from proceedings of seminars and conferences, theses and dissertations will be projected, including the percentage of content taken from various institutional repositories, including the digital library content of that institution as well. In brief, the projected percentage of similar content will as here below.

1. Overview in percentage
2. Every source projected with percentage
3. Breakdown giving each sources
4. Comparison of manuscript against already published sources, making provision for side-by-side comparison of similar content and plagiarized content.

5. Summary of the report of similar content and provision for its download.

**Repercussions of Plagiarism**

Plagiarism is punishable offence. On account of the growing population of scientists and researchers and the pressure to publish, an increasing number of scientists and researchers have taken to cutting corners, resulting in falsified research, fraudulent data, paraphrasing, duplication and blatant plagiarism. The entire undesirable attempts is classified under 'research misconduct'. This research misconduct creates a ripple effect of costly damages for researchers, organizations and the general public as well, including the impact on patients to a greater extent. Organizations may miss financial support on account of the frauds and misconduct and their undesirable impact on the public and society. The boundary between plagiarism and quality research work is unclear. However, it is essential to overcome plagiarism; avoiding plagiarized content in research reporting is an essential requisite. The software supports fighting against plagiarism by projecting similar ideas drawn from published sources. It is also the fact that there is high risk of being blacklisted by the search engines. The repercussions of plagiarism are

1. Immediate failure of research
2. Value of the research is questionable
3. Retraction of research paper
4. Black listing of scientist and institutions
5. Subject of discussion forums
6. Punishable offence and fine can be imposed in court of law

The study of some of the prominent cases of plagiarism over the past five years have shown how does plagiarism seriously affects individuals, institutions and public at large. The case of University of Bayreuth - the German Defense Minister proved how even the mighty can fall in the wake of research misconduct. People normally remain silent in case of ordinary research scholar. However, when he raises to power and assumes office of power and authority, immediately they start verifying/downloading his or her thesis and research publications. Intention is mainly to prove that there is no quality in the research output, rather to a greater extent to project unethical research misconduct. Individual costs of plagiarism when it is proved results in the

1. Loss of jobs
2. Revoked Ph.D -doctoral degrees
3. Lawsuits filed for compensation
4. Questioned integrity.

Duke University also underwent a case of alleged misconduct by one of the cancer researcher. For five years Duke University backed research and medical treatments over a hundreds of cancer patients. The research findings and recommendations were published in the most prestigious medical journals, including Nature Medicine, The Lancet Oncology, Blood
and PLoS ONE. After investigation in 2010, researcher was accused of falsifying his research data and committing fraud. Finally, Duke University faced with the challenge of recovering from a damaged reputation, amongst the medical community, publishers and patients worldwide.

The Journal of Medical Ethics published a survey research in 2011 examining 180 retracted papers which contained research linked to 28,000 patients. The study found that 6,573 of those patients were treated utilizing the data and research from the retracted papers.

The New York Quarterly Magazine’s Editor, Raymond Hammond, describes his experience of the facing the practical incident of plagiarism which really caused damage. He had to rush to withdraw the affected issue of the magazine from sale, subsequently he lost considerable amount of money for being unable to circulate that issue for sometime, moreover, he had to meet the advocates regularly and plan a course of action that would get relief and finally, make provision to sell the issue of the magazine again. The Hindu newspaper withdrew recently the article published by Sri Veerappa Moiley on account of detection of similar content in the paper published. Long-term cost associated with research misconduct is tarnishing the brand name. Academic and research institutions, universities, publishers and journals -all have “brands.” They have invested to foster their authority and reputation, towards attracting more students, selling more of their publications and also to get funds from government agencies. In an interview with Dr. Benson Honig, says, “I am well aware of the universities that have found scandalous things, for example, a professor taking student work and publishing it as his own... And they kept a lid on this... because researcher is an employee, and the case of misconduct involving that researcher has the potential to harm the institution's brand”.

Money spent to investigate the case of plagiarism and research misconduct, as well as, costs associated with repairing the damage of reputation of the organizations is embarrassing and sensitive - involving lawsuits, legal fee, investigation costs, loss of grant sanctioned. There are instances of spending to a tune of $525,000 for investigating into the research misconduct. The process involves formation of a committee of experts, payment of sitting fee, flight charges and so forth. Most of the cases, it is kept confidential as it is embarrassing.

The iParadigm, California (Turnitin) has conducted a survey covering 879 higher and secondary educators from around the world to project the types of plagiarism.

1. CLONE: Submitting someone’s work word-for-word as one’s own.
2. CTRL-C: Contains significant portions of text taken from a single source
3. FIND-REPLACE: Changing key-words and phrases but retaining the same ideas
4. REMIX: Paraphrasing content from many sources and present in a sequence
5. RECYCLE: Borrow from the author’s previously published sources without citation
6. MASHUP: Mixing textual contents borrowed from multiple sources
Avoiding Plagiarism in Doctoral Theses

Publishers of scholarly journals insist on plagiarism check before accepting the research papers for publication. They in fact fix tolerance limit or restrict the percentage of similar content. Some of the reputed publishers, of late, are rejecting 23% of the submissions on account of the plagiarism or self-plagiarism to an enormous extent. In case of research papers of journals, very little of similar content is expected, especially in the paragraphs covering introduction/background of the topic, definitions, objectives and methodology. The rest of the content will be usually nothing but interpretation and inferences drawn from the original data. Of course, findings of the research, discussions, recommendations and conclusions, there cannot be similar content unless the author cites similar studies conducted previously. However, this is not so in case of doctoral thesis. Drafting doctoral thesis involves survey of literature over a period of 2 to 5 decades. Further, many of the definitions and important studies conducted so far in the discipline have to be discussed in detail. Sometimes, about the organization selected for the study has to be explained, covering the aspects of establishment and developments, including products and services offered. While tracing such of these issues, one has to rely upon previously published sources, and hence the percentage of similar content will obviously increase. Percentage of similar content will increase:

1. In case of descriptive research
2. While tracing origin of the concept or historical background
3. In case of narrating developmental aspects
4. While projecting views of several authors and researchers.
5. In case of subjects like philosophy, religion, sociology, literature, biography and so forth where the research relies much on existing literature and old classics.

In most of the above cases, the percentage of similar content will be high. Hence, the tolerance limit may be allowed to an higher extent in social science and humanities than science and technology. Above all it is an accepted fact that Plagiarism Detection Tools/Software provides opportunity for the Research Scholar and the Research Supervisor to improve the originality of research output. The merits and usefulness of the tool are predominantly to

1. Improve originality of research contribution.
2. Correct the statements/quotes in case of inaccuracies or omissions.
3. Rectify the citations when they are incomplete
4. Avoid plagiarism while referring to similar content depicted
5. Trace the original information sources, contents and authors
6. Reduce percentage of similar content in case it is very high.

Popular subject expert in the area of management science, Dr. Benson Honig recommends to impose standards on intellectual property as an important part of a solution: “First and foremost,
I don't think any article should be accepted, at a conference or for journal, until the author has voluntarily put it through a plagiarism checking system and verified that it's clean.

If you purchase a car, you have to verify the titles. If you purchase a house, you have to go through a title deed. We need to verify that the intellectual property is truly owned before we try to publish and sell it.

Referencing style has to be in a standard format. An accepted standard format has to be adhered to. Otherwise, the software does not exclude the quoted text. Improper citations, incomplete quotes, wrong quotes, absence of original author and source leads to the increase in the percentage of similar content. It is important to note that whatever bibliographic content is copied and pasted, the software treats it as plagiarized content, eventually leading to an increase in percentage of similar content. Hence, it is suggested to use 'end note' or 'zotaro' or any other chosen software, for that matter one can type out the bibliographic details as well. Here, the researchers have to adhere to a standard format for rendering references and bibliography. Further, uniformity and consistency should be maintained while rendering references and listing bibliography.

'Bibliography' is nothing but list of sources consulted. However, on the other hand, 'References' points to the list of sources actually cited or referred in support of the researcher's new findings. Every scholarly journal has got its own style sheet. Abide strictly to a journal's request for citation formats such as APA (American Psychological Association Standard), MLA (Modern Language Association) or AMS (American Mathematical Association). You can find instructions for and examples of each reference formatting style online.

Conclusion

Recently, the percentage of plagiarism in research is increasing. The UGC will forward complaints to the concerned authorities of the University for initiating action against the researcher and the supervisor. The university has to formulate the committee comprising of experts to look into the matter. In case the guide and the researcher are found guilty, the research content will be withdrawn without any correspondence. Anti-plagiarism tool is not for policing but, it will help the researcher and the guide to improve the originality of research output. The tools will be of use to the researchers in tracing the original quotes and information sources. Thereby, one can rectify citations as well as the quotes in case of mistakes. The tool avoids students copying the assignments of classmates. Inculcates the strong habit of creating ideas based on reflective thinking. Also helps in developing writing skills. In this way, it really helps the researchers and students in developing creative ideas, writing skills and a good command over the language. It is essential to acknowledge authors or cite the sources from where ideas or sentences are borrowed. While rendering references, a standard format has to be adopted. While purchasing a motor car, buyer usually verifies the documents against malefic intentions. Again, while purchasing a house, people have to go through a title deed. On similar lines, one need to verify research contents to clarify that the intellectual property is truly owned before one tries to publish and sell it.
Useful Open Access Resources in Physics

M.K. Bhandi, Nagaraj, M.N.  
1 University Librarian, Mangalore University, Mangalore  
2 Assistant Librarian, Raman Research Institute, Bangalore

Abstract

Open access (OA) publishing is substitute model for the publication of scholarly journals. It makes articles freely available to readers on the internet and covers the costs associated with publication through means other than subscriptions. The main aim of the study was to identify open access resources and to study their actual usefulness in the physics research. This paper discusses different types of open access resources used by physics researchers. The researcher visited many Physics related websites, and OA forums/blogs for locating open access resources. The enumerate list is prepared focusing mainly on open access journals, Institutional repositories (IR), preprints repository like arXiv, NASA Astrophysics Data Systems (ADS), open access courseware and are discussed along with findings.