

To Assess the Effectiveness of Anti-Plagiarism Tools



Payal B. Dahotre, Shobha K. Bawiskar

Abstract: Plagiarism is “Someone using someone else’s intellectual product”. Now a day’s Plagiarism is increasing which violates an Intellectual property right which is serious cyber crime. Everyone is now becoming aware of these types violations. So checking the Plagiarism by using anti plagiarism tools is very important task. In this research article textual based plagiarized data samples of 1 GB was created and by using various free-ware software’s and License software and by considering various parameters its efficiency assessment is studied. As a result of this study everyone in society will be able to detect plagiarisms which will avoid Intellectual property right related crimes.

Keywords: Plagiarism, Anti-plagiarism software, Urkund, Plagiarism Checker x, Plagiarism Detector

I. INTRODUCTION

Plagiarism is a common problem that is often the result of lack of knowledge and skills. Plagiarism is an act of fraud it involves both the stealing someone else’s work and lying about it afterward. People do not know much about plagiarism.. People lack knowledge about proper citations. Plagiarism could be avoided if all people who writes articles and research paper know when and how to cite the ideas and text they had taken from other authors.[4]People who rush to do things on the last minute are meant to do the last things that would make their tasks easier-plagiarize. People find the internet a good medium for searching and copying related topics. People copying the works of other people do not believe the original creators would catch them. Plagiarism can be detected by using some plagiarism tools, software’s like Plagiarism checker x, plagiarisma, plag tracker, plagscan, copy leaks etc.[6]. Every plagiarism checker tool has some unique features. It has its advantages and disadvantages. There is a wide selection to choose from depending on one’s needs. It is Advisable to check out the different plagiarism checkers before settling down for one. Some of the free tools have their limitations and the paid services are often a better choice. [1-3]

Objectives:

- To create the textual based plagiarized data samples of 1 GB with varying plagiarized percentage.
- To give input using two methodology One by copy and pasting and another is Summarizing methodology
- To test the various parameters of anti-plagiarism software by comparison.
- To check the authenticity & limitation of software.

Algorithm:

- Step No 1: start
- Step No 2: create the data plagiarized data samples with Advanced Plagiarised Data in (%) Category by using various parameters and methodologies (cut and paste methodology is selected)
- Step No 3: Demonstrate the samples with software tools selected and do the observations
- Step No 4: comparative analysis of results obtained
- Step No 5: Stop

Methodology Used: Experimental approach is used.

For Input-

Data Sample Creation- 1 GB data sample was created by using some research paper data.. The main criteria used to made samples is Copy paste method. From research paper data selected was copy pasted together for making Single paragraph of textual [5] based samples. Totally 35 samples were made - each sample having various known percentage of plagiarism data is inserted advanced while creating the samples.

For Experimental Processing

2. The plagiarized data samples created were given input to software’s [6] on which observation and testing will be done. For Experimental method there are n-tools available online like Plagiarism checker x, plagiarisma, plag tracker, plagscan, copy leaks etc. but selection of open source freeware Anti-plagiarism tools[4] used for experimental purposes was made based on most widely rated and used software’s who cannot have access to licesenced Anti-plagiarism tools which are quite expensive in terms of prize. For this work following freeware and Licensed software’s are used[2]

- a) **Free-ware software’s are:**
- i. Plagiarism checker x
 - ii. Plagiarism Detector

Manuscript received on March 15, 2020.

Revised Manuscript received on March 24, 2020.

Manuscript published on March 30, 2020.

* Correspondence Author

Payal B. Dahotre*, Assistant Professor, Yashwantrao Chavan Institute of Science, (Autonomous) Satara.

Dr. Shobha K. Bawiskar, Assistant Professor, Government Institute of Forensic Science. Aurangabad.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

To Assess the Effectiveness of Anti-Plagiarism Tools

b) **License Software:** Urkund

Observation and Result:

Following parameter is taken into account:-

- i. Large content
- ii. Matching limit
- iii. Case sensitive
- iv. Clean Invalid spaces
- v. Keyword usage

Techniques Used to Plagiarized Sample are- Cut Paste and Summarized method as shown in Table No.1 of which only cut and paste samples are selected for demonstration

purpose. and its class category - Advanced Plagiarised Data Sample in (%) is shown in Table No 2 in which we already know the Plagiarisms amount well in advanced.

Table No: 3, Table No: 4 and Table No: 5 shows the observations made by the free ware software tools Plagiarism checker x, Plagiarism Detector and License software tool Urkund respectively. And Table No 6 shows the final results of our observations based on it the conclusions are made.

Table No.1: Parameter used to make Plagiarised Samples

Sample No	Advanced known Plagiarized Data in (%)	Technique Used to Plagiarized Sample	Parameter Used					Result
			Large Content	Matching Limit	Case Sensitive	Clean Invalid Spaces	Keyword Usage	
1.	50%	Cut Paste	✓	✓	✓	✓	✓	50%
2.	30%	Cut Paste	✓	✓	✓	✓	✓	30%
3.	63%	Cut Paste	✓	✓	✓	✓	✓	63%
4.	80%	Cut Paste	✓	✓	✓	✓	✓	80%
5.	77%	Cut Paste	✓	✓	✓	✓	✓	77%
6.	00%	Cut Paste	✓	✓	✓	✓	✓	00%
7.	20%	Cut Paste	✓	✓	✓	✓	✓	20%
8.	00%	Cut Paste	✓	✓	✓	✓	✓	00%
9.	06%	Cut Paste	✓	✓	✓	✓	✓	06%
10.	31%	Cut Paste	✓	✓	✓	✓	✓	31%
11.	00%	Cut Paste	✓	✓	✓	✓	✓	00%
12.	00%	Cut Paste	✓	✓	✓	✓	✓	00%
13.	00%	Cut Paste	✓	✓	✓	✓	✓	00%
14.	00%	Cut Paste	✓	✓	✓	✓	✓	00%
15.	100%	Cut Paste	✓	✓	✓	✓	✓	100%
16.	90%	Cut Paste	✓	✓	✓	✓	✓	90%
17.	30%	Cut Paste	✓	✓	✓	✓	✓	30%
18.	30%	Cut Paste	✓	✓	✓	✓	✓	30%
19.	26%	Cut Paste	✓	✓	✓	✓	✓	26%
20.	77%	Cut Paste	✓	✓	✓	✓	✓	77%
21.	72%	Cut Paste	✓	✓	✓	✓	✓	72%
22.	50%	Cut Paste	✓	✓	✓	✓	✓	50%
23.	85%	Cut Paste	✓	✓	✓	✓	✓	85%
24.	10%	Cut Paste	✓	✓	✓	✓	✓	10%
25.	13%	Cut Paste	✓	✓	✓	✓	✓	13%
26.	00%	Summarized	✓	✓	✓	✓	✓	00%
27.	00%	Summarized	✓	✓	✓	✓	✓	00%
28.	00%	Summarized	✓	✓	✓	✓	✓	00%
29.	00%	Summarized	✓	✓	✓	✓	✓	00%
30.	00%	Summarized	✓	✓	✓	✓	✓	00%
31.	00%	Summarized	✓	✓	✓	✓	✓	00%
32.	00%	Summarized	✓	✓	✓	✓	✓	00%
33.	00%	Summarized	✓	✓	✓	✓	✓	00%
34.	00%	Summarized	✓	✓	✓	✓	✓	00%
35.	00%	Summarized	✓	✓	✓	✓	✓	00%

Table no.1 shows that the methods used to plagiarise samples like cut paste method, summarized method, Paraphrasing method. In the current research paper to making the plagiarised sample cut paste & summarized method used to Plagiarized sample. Whereas for cut paste

method and summarised method, same topic of paid research paper were selected and for making samples. Samples was prepared in which sample plagiarism values are known.

Table No.2: Advanced Plagiarised Data Sample in (%) made during sample creation.

Sample No	Advanced Plagiarised Data in (%) Category	Technique Used to Plagiarized Sample	Parameter Used				
			Large Content	Matching Limit	Case Sensitive	Clean Invalid Spaces	Keyword Usage
6,8,9,11, 12,13,14,24	00%-10%	Cut Paste	✓	✓	✓	✓	✓
7,25	11%-20%	Cut Paste	✓	✓	✓	✓	✓

2,17, 18,19	21%-30%	Cut Paste	✓	✓	✓	✓	✓
10	31%-40%	Cut Paste	✓	✓	✓	✓	✓
1,22	41%-50%	Cut Paste	✓	✓	✓	✓	✓
-	51%-60%	Cut Paste	✓	✓	✓	✓	✓
3	61%-70%	Cut Paste	✓	✓	✓	✓	✓
4,5,20,21	71%-80%	Cut Paste	✓	✓	✓	✓	✓
23,16	81%-90%	Cut Paste	✓	✓	✓	✓	✓
15	91%-100%	Cut Paste	✓	✓	✓	✓	✓

Table No.2 showing that samples which were made having value of Plagiarised sample in advance in percentage and method used to plagiarised the sample is cut paste method, with the help of parameter Large content, matching limit, case sensitive, clean invalid spaces, keyword usage Samples were detected.

Table. No.3: Comparative table showing the Difference between Self made plagiarised Sample result in (%) with Tool Plagiarism Checker X Result in (%) Duplicated

Sample No	Advanced Plagiarised Data in (%) Category	Technique Used to Plagiarised Sample	Parameter Used					Advanced known Plagiarized Data in (%)	Tool Plagiarism Checker X Result in % Duplicated
			Large Content	Matching Limit	Case Sensitive	Clean Invalid Spaces	Keyword Usage		
6	00%-10%	Cut Paste	✓	✓	✓	✓	✓	00%	99%
8			✓	✓	✓	✓	✓	00%	90%
11			✓	✓	✓	✓	✓	00%	81%
12			✓	✓	✓	✓	✓	00%	99%
13			✓	✓	✓	✓	✓	00%	94%
14			✓	✓	✓	✓	✓	00%	99%
24			✓	✓	✓	✓	✓	10%	61%
25	11%-20%		✓	✓	✓	✓	✓	13%	73%
7	21%-30%		✓	✓	✓	✓	✓	21%	79%
19			✓	✓	✓	✓	✓	26%	100%
2			✓	✓	✓	✓	✓	30%	76%
9			✓	✓	✓	✓	✓	6%	90%
17			✓	✓	✓	✓	✓	30%	94%
18			✓	✓	✓	✓	✓	30%	100%
10			31%-40%	✓	✓	✓	✓	✓	06%
1	41%-50%	✓	✓	✓	✓	✓	50%	99%	
22		✓	✓	✓	✓	✓	50%	58%	
-	51%-60%	✓	✓	✓	✓	✓	----	----	
3	61%-70%	✓	✓	✓	✓	✓	63%	100%	
4	71%-80%	✓	✓	✓	✓	✓	80%	100%	
5		✓	✓	✓	✓	✓	77%	100%	
20		✓	✓	✓	✓	✓	77%	59%	
21		✓	✓	✓	✓	✓	72%	97%	
23	81%-90%	✓	✓	✓	✓	✓	85%	43%	
16		✓	✓	✓	✓	✓	90%	94%	
15	91%-100%	✓	✓	✓	✓	✓	100%	100%	

Table No.3 showing that Range of Advanced Plagiarised samples in Percentage, with the help of cut paste method and Parameters were applied..Sample analysis done by using Plagiarism Checker X software which showing above result value in Percentage (%). It also showing that the difference between the Advanced known Plagiarized Data in Percentage (%) and result value in percentage (%) given by Plagiarism Checker x Tool.

Table. No. 4: Comparative table showing the Difference between Self made plagiarised Sample result in (%) with Tool-Plagiarism Detector Result in (%) Duplicate (D), Original (*O), Plagiarised (P)

Sample No	Advanced Plagiarised Data in (%) Category	Technique Used to Plagiarised Sample	Parameter Used					Advanced known Plagiarized Data in (%)	Tool-Plagiarism Detector Result in (%) Duplicate(D), Original (*O), Plagiarised(P)
			Large Content	Matching Limit	Case Sensitive	Clean Invalid Spaces	Keyword Usage		
6			✓	✓	✓	✓	✓	00%	94% D
8			✓	✓	✓	✓	✓	00%	100% O
11			✓	✓	✓	✓	✓	00%	94% P

To Assess the Effectiveness of Anti-Plagiarism Tools

12	00%-10%	Cut Paste	✓	✓	✓	✓	✓	00%	88% P
13			✓	✓	✓	✓	✓	00%	96% P
14			✓	✓	✓	✓	✓	00%	89% D
24			✓	✓	✓	✓	✓	10%	90% P
25	11%-20%		✓	✓	✓	✓	✓	13%	90% P
7	21%-30%		✓	✓	✓	✓	✓	21%	94% O
19			✓	✓	✓	✓	✓	26%	00% P
2			✓	✓	✓	✓	✓	30%	97.12% D
9			✓	✓	✓	✓	✓	6%	99% D
17			✓	✓	✓	✓	✓	30%	95% P
18			✓	✓	✓	✓	✓	30%	74% P
10	31%-40%		✓	✓	✓	✓	✓	06%	94% D
1	41%-50%		✓	✓	✓	✓	✓	50%	100% O
22			✓	✓	✓	✓	✓	50%	36% P
-	51%-60%		✓	✓	✓	✓	✓	----	----
3	61%-70%		✓	✓	✓	✓	✓	63%	83% D
4	71%-80%		✓	✓	✓	✓	✓	80%	100%
5			✓	✓	✓	✓	✓	77%	98.06% D
20			✓	✓	✓	✓	✓	77%	03% P
21			✓	✓	✓	✓	✓	72%	27% P
23	81%-90%		✓	✓	✓	✓	✓	85%	94% P
16			✓	✓	✓	✓	✓	90%	95% P
15	91%-100%		✓	✓	✓	✓	✓	100%	90% P

Table No.4 showing that Range of Advanced Plagiarised samples in Percentage, with the help of cut paste method and Parameters were applied. Sample analysis done by using Plagiarism Detector software which showing above result

value in Percentage (%). It also showing that the difference between the Advanced known Plagiarized Data in Percentage (%) and result value in percentage (%) given by Plagiarism Detector Tool.

Table. No .5: Comparative table showing the Difference between Self made plagiarised Sample result in (%) with Tool Urkund Result in (%) Significance

Sample No	Advanced Plagiarised Data in (%) Category	Technique Used to Plagiarised Sample	Parameter Used					Advanced known Plagiarized Data in (%)	Tool Urkund Result in (%) Significance
			Large Content	Advanced known Plagiarized Data in (%)	Case Sensitive	Clean Invalid Spaces	Keyword Usage		
6	00%-10%	Cut Paste	✓	✓	✓	✓	✓	00%	70% Significance
8			✓	✓	✓	✓	✓	00%	77% Significance
11			✓	✓	✓	✓	✓	00%	00% Significance
12			✓	✓	✓	✓	✓	00%	100% Significance
13			✓	✓	✓	✓	✓	00%	No Significance
14			✓	✓	✓	✓	✓	00%	99% Significance
24			✓	✓	✓	✓	✓	10%	07% Significance
25	11%-20%		✓	✓	✓	✓	✓	13%	91% Significance
7	21%-30%		✓	✓	✓	✓	✓	21%	52% Significance
19			✓	✓	✓	✓	✓	26%	83% Significance
2			✓	✓	✓	✓	✓	30%	95% Significance
9			✓	✓	✓	✓	✓	6%	74% Significance
17			✓	✓	✓	✓	✓	30%	67% Significance
18			✓	✓	✓	✓	✓	30%	100% Significance
10	31%-40%		✓	✓	✓	✓	✓	06%	82% Significance
1	41%-50%		✓	✓	✓	✓	✓	50%	97% Significance
22			✓	✓	✓	✓	✓	50%	46% Significance
-	51%-60%		✓	✓	✓	✓	✓	----	----
3	61%-70%		✓	✓	✓	✓	✓	63%	92% Significance
4	71%-80%		✓	✓	✓	✓	✓	80%	63% Significance
5			✓	✓	✓	✓	✓	77%	47% Significance
20			✓	✓	✓	✓	✓	77%	34% Significance
21			✓	✓	✓	✓	✓	72%	54% Significance
23	81%-90%		✓	✓	✓	✓	✓	85%	69% Significance
16			✓	✓	✓	✓	✓	90%	32% Significance
15	91%-100%	✓	✓	✓	✓	✓	100%	54% Significance	

Table No.5 showing that Range of Advanced Plagiarised samples in Percentage, with the help of cut paste method and Parameters were applied. Sample analysis done by using Urkund software which showing above result value in

Percentage (%). It also showing that the difference between the Advanced known Plagiarized Data in Percentage (%) and result value in percentage (%) given by Urkund Tool.

Result: Major Findings shown in Table No: 6: as follows

Table No: 6-Comparative Analysis of all 25 samples

Sample No	Advanced Plagiarised Data in (%) Category	Technique Used to Plagiarised Sample	Advanced known Plagiarized Data in (%)	Tool Plagiarism Checker X	Tool-Plagiarism Detector	Tool Urkund	
				Result in % Duplicated	Result in (%) Duplicate(D), Original (*O), Plagiarised(P)	Result in (%) Significance	
6	00%-10%	Cut Paste	00%	99%	94% D	70% Significance	
8			00%	90%	100% O	77% Significance	
11			00%	81%	94% P	00% Significance	
12			00%	99%	88% P	100% Significance	
13			00%	94%	96% P	No Significance	
14			00%	99%	89% D	99% Significance	
24			10%	61%	90% P	07% Significance	
25	11%-20%		13%	73%	90% P	91% Significance	
7			21%	79%	94% O	52% Significance	
19	21%-30%		26%	100%	00% P	83% Significance	
2			30%	76%	97.12% D	95% Significance	
9			6%	90%	99% D	74% Significance	
17			30%	94%	95% P	67% Significance	
18			30%	100%	74% P	100% Significance	
10			31%-40%	06%	78%	94% D	82% Significance
1			41%-50%	50%	99%	100% O	97% Significance
22			50%	58%	36% P	46% Significance	
-	51%-60%		----	----	----	----	
3	61%-70%		63%	100%	83% D	92% Significance	
5	71%-80%		77%	100%	98.06% D	47% Significance	
20			77%	59%	03% P	34% Significance	
21			72%	97%	27% P	54% Significance	
23			81%-90%	85%	43%	94% P	69% Significance
16	91%-100%		90%	94%	95% P	32% Significance	
15			100%	100%	90% P	54% Significance	
Average				82.52	76.80	60.88	

Table No.6 Showing that comparative analysis of Samples among the Softwares used for detection. This table represents the known value of Original Plagiarised Samples with their range in Percentage .By analysing the Plagiarized samples with the help of Plagiarism Software including Open source and License ware software the value were getting from tools given in the table in the form of Percentage (%).

II. CONCLUSION:

For this research, nearly twenty plagiarism software [6] tools are taken for study. By studying, analyzing feature three software’s The Plagiarism checker x, Plagiarism detector [6], Urkund were selected and taken for doing the work of content analysis and its comparison was done.[5] .Plagiarism Detector gives more authenticates results as its results average value is closely resembling to Urkund i.e. As Compared to Plagiarism Checker X if free ware tools are considered. Hence it can be concluded Plagiarism Detector gives better results that and as Urkund is licesenced software its results obtained are good as 100% accurate results achieved.

Limitations:

During the analysis some limitations are encountered which are as follows:

1. All the content of the sample has been not analyzed successfully due to large size of samples only 150 words are considered in Plagiarism checker x and no binding in Plagiarism Detector

2. Only Textual [5] data samples are considered.
3. They can’t prove or show evidence that document has been plagiarized from which document or sources. Only shows, overview and comparison of plagiarism detection tools.
4. Free-ware software used are with trial version in this limitation of word can be given as input because of which while analyzing Parameter includes or exclude affects the results
5. According to software parameters seems to changed.
6. Only two tools used for this experiment.

Future Scope:

1. N-Number freeware software are available but due to time limitation only two freeware Plagiarism Checker X & Plagiarism Detector were used for experimental purpose rest of all the tool can be assess.
2. Work can be extended to video and audio plagiarism detection too as very negligible research work is found with reference to audio and video data samples are concern.
3. Can create a Software tool which can remove all the limitations like dynamic input size, excellent plagiarism detection [6] capacity.



To Assess the Effectiveness of Anti-Plagiarism Tools

4. Created tool should be made freely available to users who are interested in using it this is because not everyone has the access to licesenced tool as it is very costly can be seen in big organizations or educational institutions or university. Now a day's Turnitin and Urkund widely used Licensed Anti-plagiarism software

REFERENCES:

1. International Conference on Computer Systems and Technologies - CompSysTech'07Computer- Based Plagiarism Detection Methods and Tools: An Overview Romans Lukashenka, Vita Graudina, JanisGrundspenkis
2. Alzahrani, S. Salim, N. Abraham, A. Understanding Plagiarism Linguistic Patterns, Textual Features and Detection Methods, preprint 2011. 2.
3. Brin, S. and Davis, J. and Garcia-Molina, H. "Copy Detection Mechanisms for Digital Documents". In: ACM International Conference on Management of Data (SIGMOD 1995), May 22-25, 19953.
4. Younghwa Lee, Elsevier Understanding anti-plagiarism software adoption: "An extended protection motivation theory perspective 2010".
5. Michal Duracika, Emil Krsaka, Patrik Hrkuta, and Science Direct2017, "Current trends in source code analysis, plagiarism detection and issues of analysis big datasets".
6. Bart Penders, 7 July 2017, Springer.M. Ryn anen and A. Klapuri, "Query by humming of midi and audio using locality Sensitive hashing,"in Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2008.

AUTHORS PROFILE



Miss Payal Bharat Dahotre has completed B.Sc Forensic Science and M. Sc Forensic Science in Government Institute of Forensics Science Aurangabad in Department of Digital and Cyber Forensics. Recently working as Assistant Professor in Yashavantrao Chavan Institute of Science, (Autonomous) Satara, since last one and half year. Topics interests are of are cyber

investigation, incidence response, network Forensic, Multimedia Forensics, etc... further wish to apply for PhD too.



Dr. Shobha Bawiskar has completed BCS degree and M.Sc Computer Science from College of IT & Management , Vivekanand college of Arts , science and commerce respectively affiliated to Dr.Babasaheb Marathwada University, M.Phil from YCMOU from affiliated IMTR study centre in Aurangabad and received Ph.D in Computer Science from Dr.Babasaheb

Marathwada University. 13+ years of teaching experience. Have published twenty five plus research articles in various national / international conferences, Journals, Seminars. Currently working as Assistant Professor in Government Institute of Forensic Science. Aurangabad. Topic of research interest includes Digital and Cyber forensic, Mobile Forensics, Network forensics, Nanotechnology, image processing, Internet Of Things.