

Performance evaluation of free anti-plagiarism software

Amalia

Department of Informatics Engineering, Islamic University Of North Sumatra, Medan, Indonesia. Corresponding Author: amalia@uisu.ac.id

Abstract. Plagiarism is a very despicable act in the academic world that requires an active effort from Universities to prevent and to identify it at an early stage in order to maintain the reputation of the university. One easy way to quickly identify plagiarism is by the university proactively conducting a deep examination of papers written by members of their academic community before it is published. Nowadays, there are several anti-plagiarism software that can assist to quickly conduct these plagiarism examinations. However, because of these many choices of anti-plagiarism software, even freely available software, sometime it is hard to choose the most suitable software to use. This paper report on the result from a study that compare several freely available anti-plagiarism software with the hope it can help universities on deciding which one to use. This study chose the top six highest ranked software that were produced from several search engine with the keyword "free anti plagiarism software", which are: Plagiarisma.Net, Viper, The Plagiarism Checker, Dupli Checker, Plag Tracker and Small Seo Tools Plagiarism Checker. The study employed several important parameters in comparing these anti-plagiarism software, which are: accuracy, validation time, the diversity of format text, number of characters/words, ease of use, report, and security. Results from the study indicate that there is no one anti-plagiarism software that can meet all user needs, thus there is no one superior software for all categories but there are software that are superior for certain categories.

Key words: Performance of Free Anti-Plagiarism Software, Anticipation of Plagiarism, Plagiarism in University, Comparison of Free Anti Plagiarism Software.

Introduction

Plagiarism is commonly defined as „*the act of taking or attempting to take or to use (whole or parts) of another person’s works, without referencing or citation him as the owner of this work*” (Overview and Comparison of Plagiarism 2011). Plagiarism in academia is a despicable act because it does not respect the integrity of other’s works. According to Thomas (2008), plagiarism is a pervasive form of academic dishonesty in a university setting since it distorts learning and assessment where determining and detecting it are crucial to maintaining academic integrity.

The practice of plagiarism is also a lot going on in the academic society, students, and even researchers. Since its easily search for and download the papers of others in the internet, make the practice of plagiarism is increasing. So, it requires some steps to prevent and identify at an early stage in order to maintain the good name of the university. One way to early identify plagiarism is that university proactively conduct a deep examination on papers written by the academic community before published. In addition, every academic society have to prevent their works from accidental plagiarism such as forgetting to include the citation.

One way to detect the academic writing is using Plagiarism Detection Software. Asim M. El Tahir Ali et al declared that it is better and it’s a must that all fields of academics to use plagiarism detection software to stop or to eliminate students cheating, copying and modifying documents (Overview and Comparison of Plagiarism 2011). Another research about this was done by Dr. Ann Ledwith and Angélica Rísquez which indicate that the use of anti-plagiarism software led to a decrease in Internet plagiarism and lower grades awarded in peer reviews. Additionally, students were found to have a positive attitude towards the anti-plagiarism software in the context of peer reviewed assignments(Rísquez 2010).

A variety of free and commercial software applications designed to detect plagiarism from Internet sources has appeared in recent years (Risque 2010). However, because of so many choices of anti-plagiarism software, sometimes it makes some users confused in choosing the most suitable software to use, from the baseline survey revealed that some users have to change anti-plagiarism software several times because it does not suit their needs. The previous studies have evaluated some of commercial anti-plagiarism software that have been widely used in top universities in the world. But, until now there are still no specific discussion about the free Plagiarism Detection Software (PDS). Thus, this paper aims to compare several free anti-plagiarism software. It can be used as reference in selecting the best free anti-plagiarism software to use.

Materials and Methods

Selecting Anti-Plagiarism Software To Evaluated

Of the many available free anti-plagiarism software, this study chose some software that have the highest ranks (first page) in the search results on some search engines with the keywords "free anti plagiarism software". Even though, some of them are excluded from the list since they are only available for free trial for several days. And then one software which is evaluated before, i.e. Plagiarism Detect at www.plagiarsm-detect.com had to be removed from the list because the software developer has confirmed that they did not continue their free services. Some software that provide free services, but still offer an upgrade to a pro, are still kept since there are still some parameters which can be evaluated. The evaluated software are:

1. Plagiarisma.Net (www.plagiarisma.net)
2. Viper (www.scanmyessay.com)
3. The Plagiarsm Checker (www.dustball.com/cs/plagiarism.checker)
4. Dupli Checker (www.duplichecker.com)
5. Plag Tracker(www.plagtracker.com)
6. Small Seo Tools Plagiarism Checker (www.smallseotools.com/plagiarism-checker)

How It Works

Generally, the Plagiarism detection software works by extracting the text from the document being checked, and comparing the text with other sources like from a customized database, or from the Internet, or both (Heather 2010). According to information from official sites, the evaluated software we've found are Plagiarisma.net, Viper and Plag Tracker have their own databases containing billion of papers. So they work by comparing the similar phrases from both, i.e. from the internet and their databases. The Plagiarism Checker, Dupli Checker and Plagiarism Checker work by comparing similar phrases with the world wide web, but we did not get information whether they have their own paper's database or not. If the software found similar phrases it will give result as potential plagiarism.

Essay Specimen To Trial Test

Initial inspiration for the determination of our essay specimens is from the paper of Dr. Joanna Bull, et al, entitled Technical Review of Plagiarism Detection Software Report. In this study we made ten essays from five majors i.e. computer, health, education, economic, and engineering. Like what was done by Dr. Joanna Bull, we also partitioned these essays into four groups, such as:

- Essays obtained from on-line essay banks.
 - Essays containing material copied and pasted from the World Wide Web.
 - Essays written in collusion with others but with no Internet material included.
 - One 'hybrid', an essay written in collusion plus containing some copied Internet material.
- We labeled the essay to make it easier to monitor.

Table 1. Essay from Paper Bank

Topic	Essay	Amount of letters	Amount of Copied	Source
Computer	1	5671	100%	http://dl.acm.org
Health	2	6237	100%	http://ehe.pitt.edu
Economic	3	7041	100%	http://journal.ui.ac.id/

Table 2. Cut and Paste From WWW

Topic	Essay	Amount of letters	Amount of Copied	Source
Computer	4	1161	58%	3 website
Economic	5	968	85%	5 website
Education	6	1318	85%	Hybrid of 2 website + 1 collusion with essay 7
Education	7	651	70%	Hybrid of 2 website + 1 collusion with 6

Table 3. Collusion Between Writers

Topic	Essay	Amount of letters	Amount of Copied	Source
Engineering	8	2852	100%	Essay No 9 and essay 10
Engineering	9	2852	100%	Essay No 8 and 10
Engineering	10	1335	100%	Essay No 8 and 9

Evaluation Parameter

Accuracy Evaluation

Accuracy is the most important thing in the PDS. To evaluate the accuracy of each software we tested them based on the available essay specimen. As mentioned earlier, every essay specimen containing some of plagiarism percent taken from various sources. Software with good accuracy is the software which returns the amount of percentages close to plagiarism percentages, and sources are shown also in accordance with the sources cited by the essay specimen. But not all of the PDS give the result in the possible plagiarism percentages. That two software are The Plagiarism Checker and Dupli Checker only showed the sources suspected as the quoted sources. Because of not generate same result formation with other software result so we exclude these software from this parameter discussion. In this study, every scan result from the essays are saved in the table 4 to be analyzed then.

For essay 1, 2 dan 3 the essay source is taken from the paper banks, the amount of plagiarism material are 100%. There is no certainty amount of text similarity percentage considered as plagiarism. According to Viper, a document considered has no plagiarism suspected or low risk plagiarism if have similarity with others source under 12%. A document considered contains high risk plagiarized material if have similarity to others source above 21% (Viper 2012). The result of these software showed none of these software give the correct result which is 100% plagiarism. The best result was produced by Plagiarisma.Net that gave a score of 79% for essay 3, 64% for essay 2, and 9% for essay 1 (low risk of being plagiarized). The worst result was produced by Plag Tracker and Viper that all gave a score of below 12% for two essays. While Small Seo Tools Plagiarism Checker is between these two extremes that produced a score between 15% - 27%, which can be categorized as medium to high risk of being plagiarized.

Table 4. Amount of Content or Percentage Text Similarity

Software	Report Result (Percentage Content That Have Similarity /Possible Plagiarism)				Amount Of Plagiarism Supposed To be
	Plagiarisma. Net	Viper	Plag Tracker	SmallSeoTools Plagiarism Checker	
No Essay					
1	9%	7%	0%	21%	100%
2	64%	10%	47%	15%	100%
3	79%	55%	8%	27%	100%
4	54%	60%	77%	42%	58%
5	44%	71%	26%	69%	85%
6	74%	70%	25%	83%	85%
7	58%	72%	67%	74%	70%
8	1%	1%	0%	3%	100%
9	0%	1%	0%	2%	100%
10	0%	0%	0%	2%	100%

For essay 4 and essay 5, which area II plagiarized texts from several websites, all of the software produced a score above 21% thus can be categorized as having a high risk of being plagiarized. For essay 6 and essay 7, which are hybrids, all of the software also produced a score above 21%. Several software even gave almost identical scores with the plagiarism essay specimen but regretfully it only detected that the plagiarized materials came from the web and not as a result from collusion. None of the software can detect collusion plagiarism.

For essay 8 through 10, which were plagiarized by colluding among itself, all six software were not able to detect this. Another main drawback from all of these software is that it cannot check with the document's references, all sentences that have similarities with its source will be categorized as having a possibility of being plagiarized. Several software do offer a pay service to accomplish this.

Validation Time Evaluation

Evaluation time is also another important parameter. Some PDS consume a lot of validation time but there're also only need a little time. The more papers to be compared, the longer validation time they used, and the more sources to be compared will make the validation time longer.

Out of the six software that were compared, The Plagiarism Checker produced the quickest result where each sentence is checked instantaneously without delay. While Plag Tracker produced a quick result for the first document but for subsequent documents there will be a long delay for the result to be produced, where the status is "In Progress" and while in this state no other document can be checked. For the other four software (Viper, Small Seo Tools Plagiarism Checker, Plagiarisma.Net, and Dupli Checker) all required almost identical time that can be correlated with the number of words contained in the document; but Plagiarisma.Net and Dupli Checker where given a low score for this test because both software did not display its stage progress so users don't have information how long the process will take.

Diversity Of Text Format

The diversity of text format which can be validated is a decisive parameter for some people. Given that in the evaluation of PDS is the free PDS, sometimes there are limitations given by the software development company. From these six software, only Plagiarisma.Net, Viper and Dupli Checker that are able to validated the documents with various text format. Plagiarisma.Net is able to validated documents with most types of text format, i.e. pdf, doc, docx, rtf, odt, txt and html. Viper is able to validated the document with the text format

doc, docx, txt, rtf, and pdf. While Dupli Checker is only able to validated documents with the text format docx and txt. For Plag Tracker-version free account, The Plagiarism Checker and Plagiarism Checker are not able to checked the essay plagiarism with some text format since the software's validation method was done by cut and paste the plain text from tje documents or essay which is validated into a textbox on the software's menu.

Number Of Character/Essay per Search

From the six software to be evaluated, only 2 of them, i.e. The plagiarism Checker and Small Seo Tools Plagiarism Checker that not bounded the amount of characters in an essay and the amount of essay to be processed in one day per account user either. While Plagiarisma.Net not bounded the amount of characters in an essay to be validated, but it's only allowed 5 essays search in a day for a registered user (Plagiarisma.Net 2013). Viper bounded the amount of scanning per user's account for 540 queries, Viper's not bounded the amount of documents to be processed if they're still at the lower limit that viper allowed (Viper 2012). Dupli Checker only allowed 1500 words or 50 Kb document's size per search with the amount of essays to be processed are maximum 50 essays for registered user (DupliChecker 2012). Plag Tracker allowed 5000 words per search but there's no limitation on the amount of documents (PlagTracker 2012).

Ease of Use

These parameters are so subjective. High or low rating for this parameter depends on the user's needs. Some PDS need installation and registration the user account to get the software's services. For users who only occasionally need an anti-plagiarism software, installation steps and registration of course will take time. But for those users who regularly use anti-plagiarism service such as teachers, lectures, librarians, etc this is not a problem, because once we've done installation and registered an account, we simply select the file or files you want to check for plagiarism and the software will generate the result.

From these six software, only Viper that need to be installed. This application is small, only 1 MB and can be downloaded from the Viper official sites. Besides having to install the application, the user also must have a Viper's account to run this service.

In addition to the above, this parameter is also evaluating ways to input essay on plagiarism checks. There are several options how to scan documents that are commonly used by the PDS such as copy-paste plain text into a the text box and upload a text file. Of the six software that provides both these features are Plagiarisma.Net and Dupli Checker. While for The Plagiarism Checker and Plag Tracker these two features are available, but for free account user they can only scanned the documents by copy-paste the plain text into the text box. For the software Small Seo Tools Plagiarism Checker only provided one feature i.e. copy-paste the plain text into a the text box. For Viper feature which can be used simply by uploading the document.

Clarity Report

Clarity Report is the proof of validation result. A good report is the report that can easily understand by users. In this study, from the six software, only Plagiarisma.Net and Viper that are able to generated the report in the form of file. For PDS which showed the matching content or their originality in percent are Viper, Plagiarisma.Net, Small Seo Tools Plagiarism Checker and Plag Tracker. while Dupli Checker and The Plagiarism Checker only showed the link content/sites which have the similarity with the validated essay, and in order to see the text similarity, users have to open the sites, but unfortunately they're not highlight their similarity. PDS Report which highlight the text that potentially plagiarism are Viper, Plag Tracker and Plagiarisma.Net. For PDS which not generated report in the form of file, the validation resulting display is a html which can be saved and re-open another time offline, except The Plagiarism Checker. The report page of the software still need to be connected to the internet to display the result of scanning plagiarism process.

Security

The evaluation for security we mean security for documents in the validation. Some issues arises that free PDS free sometimes take the scanned documents to be saved on their databases. This is a problem for some people. From the official sites of these six PDS, Plagiarisma.Net and Plag Tracker stated unequivocally that they did not save any content uploaded by user. While Viper stated that they save all the essays scanned within a secure database, their goal is to get the accurate plagiarism report from previously submitted essays. According to Viper, nobody else can access the database including lecturers, universities and other plagiarism scanners (Viper 2012). And after 9 months from the files scanned, Viper would uploaded the essay to their database and it appears on one of our study sites, so that other students may use it to help them write their own essays (Viper 2012).

For other PDS, i.e. The Plagirism Checker, Dupli Checker and Small Seo Tools Plagiarism Checker we didn't get any information about this on their official sites.

Results and Discussion

Table 5 showed the result from the evaluations overall. Plagiarism detection software is effective in intercepting plagiarized sources which elude the scrutiny of the editorial board prior to final publication. there are a lot of issues about the PDS accuracy which sometime resulting the wrong validation of the documents. All of these are depend on the algorithms that it used and also the amount of sources to be compared, such as papers databases or other resources from the internet by each PDS. Some of PDS, especially commercial ones able to access the database that contains thousands of papers from major universities to be compared. Some PDS will report the amount of percentages of writings similarity with the papers in their databases or in the internet without evaluated that the writing has a good citation. But there are also PDS which ignores the writing's similarity since they have a good citation. Off course, these two PDS will result in a different plagiarism percentage.

Table 5. Software Evaluation

Parameter	Evaluated Software Anti Plagiarism					
	Plagiarisma Net	Viper	The Plagiarism Checker	Dupli Checker	Plag Tracker	SmallSeoTools Plagiarism Checker
Accuracy						
From essay Banks	***	**	-	-	**	***
Copied paste from WWW	****	****	-	-	****	****
Collusion with Others	*	*	-	-	*	*
Hybrid	**	**	-	-	**	**
Exclude citation reference	*	*	-	-	*	*
Validation Time	**	***	****	**	**	***
Diversity Of Format Texts	****	***	*	**	*	*
Software Allowance						
Amount of Character per essay	*****	***	*****	*	**	*****
Amount of essays per a day	***	*****	*****	***	*****	*****
Ease Of Use						
Diversity of document input	*****	***	**	*****	**	**
For one scanning	****	*	****	****	*****	****
For continues Scanning	****	*****	****	***	**	***
Essay under 1500 words	****	****	****	****	****	****
Long essays	****	****	****	*	**	****
Report Clarity	***	*****	*	*	***	**
Security	****	**	*	*	****	*

The following expression denote that ***** Excellent, **** Very Good, *** Good, ** Acceptable, * Poor.

Conclusions

The main results of this study indicate from these six free PDS there is none can detect collusion plagiarism and these software generate less accurate for essay which copied from papers bank. This study also indicate none of these free anti-plagiarism software can meet all the needs of the users so that there is no the most superior software. Still and all these free PDS can be an option to academic society check plagiarism or only to prevent their works from accidental plagiarism such as forgetting to include the citation.

Acknowledgements

I would like to express my very great appreciation to Mr. Rahmad Dawood for his valuable and constructive suggestions during this research work. His willingness to give his time so generously has been very much appreciated.

References

- Bailey, Jonathan. "Plagiarism Today." *Review: Viper Anti-Plagiarism Scanner*. April 29, 2010. <http://www.plagiarismtoday.com/2010/04/29/review-viper-anti-plagiarism-scanner/> (accessed August 17, 2013).
- DupliChecker. *Dupli Checker, Free Online Software For Plagiarism Detection*. 2012. <http://www.duplichecker.com/> (accessed August 18, 2013).
- GENARO V. JAPOS, BERNARD A. GUTIERREZ, JUN BRIAN P. TUBONGBANUA. "Efficacy of Plagiarism Detection Software." *IAMURE: International Association of Multidisciplinary*, 2011.
- Heather, James. "Turnitoff: Identifying and Fixing a Hole in Current Plagiarism Detection Software." *Journal of Assessment and Evaluation in Higher Education*, 2010.
- McKeever, Lucy. "Online plagiarism detection services—saviour or scourge?" *Assessment & Evaluation in Higher Education*, Vol. 31, No. 2, April 2006: 155-165.
- "Overview and Comparison of Plagiarism." *Dateso*, 2011: 161-172.
- Plagiarisma.Net. *Plagiarisma.Net, Duplicate Content Checker For Your Essay*. 2013. www.plagiarisma.net (accessed August 17, 2013).
- Plagiarisma.net. "Turnitin have flaws as duplicate content checker." *Free Plagiarism Checker Blog-Plagiarisma.Net*. September 1, 2011. <http://blog.plagiarisma.net/> (accessed August 17, 2013).
- PlagTracker. *Plag Tracker, The Most Accurate Plagiarism Checking Service*. 2012. <http://www.plagtracker.com> (accessed August 16, 2013).
- Rísquez, Dr. Ann Ledwith and Angélica. "Using anti-plagiarism software to promote academic honesty in the context of." *Studies*, 2010.
- SmallSeoTools. *SmallSeoTools Plagiarism Checker*. 2012. <http://smallseotools.com/plagiarism-checker/> (accessed August 16, 2013).
- Thomas, Marlin. "Plagiarism Detection Software." *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (ELEARN) 2008*. Las Vegas, Nevada, USA: Curtis J. Bonk, Mimi Miyoung Lee & Tom Reynolds, 2008. 2390-2397.
- Viper. "How does Viper use my essay/dissertation?" *Viper The Anti Plagiarism Scanner*. 2012. <http://www.scanmyessay.com/viper-use-essay.php> (accessed August 17, 2013).