

Plagiarism and Detection Tools: An Overview

Urvashi Garg

Asstt. Prof., HCTM Kaithal

Abstract : *Plagiarism has always been a difficult problem to overcome. Various tools has been developed over the past few years to detect plagiarism. This paper provides an overview of plagiarism problem. The ways of reducing plagiarism is discussed. Some of the plagiarism detection tools are discussed.*

Keywords : *Plagiarism, detecting plagiarism, tools*

Introduction

Plagiarism is a serious and widespread educational issue. It is found that 64% of students knowingly copied work atleast once their studies [1]. It is reported averages of 5% of students were caught plagiarizing in a year which shows a great difference [2]. It implies a significant amount plagiarism currently undetected. No doubt 100% plagiarism can not be avoided because we want some source of information also but proper citation of source of information can avoid us from plagiarism. IEEE TV reports an increasing trend of number of plagiarism reports each year. In 2004 there were 14 cases reported to IEEE as plagiarized material whereas it increased to 50 cases in 2006 and in 2008 it reached to ore than 100 cases. IEEE categorizes different levels based on severity of plagiarism. Level 1 shows all the paper copied, level 2 with large section of someone's paper copied and level 3 in which a few section is copied [3].

Plagiarism is very dangerous to students as they are caught in cycle of plagiarism which results in their demotivating their capabilities to develop their own ideas and also their communication skills. Once the students get habituated to copy material, his creativity ability gets ruined.

Ways to reduce plagiarism

Many methods to fight against plagiarism are developed and used. These can be divided into two classes:

- (1) Methods for plagiarism prevention
- (2) Methods for plagiarism detection.

Methods for plagiarism prevention include precautionary measures. In plagiarism prevention honesty policies and/or punishment systems are framed out. Honesty policies encourage the original ideas by giving maximum incentives to students who are not involved in plagiarism.

There is a three step rule for plagiarism prevention.

All information from sources must be

1. Paraphrased, summarized or quoted
AND
2. Cited in same paragraph
AND
3. Cited again in list of references in the end of document.

Punishments are also framed out for plagiarism. Punishment depends on the severity of plagiarism.

Detecting & documenting plagiarism is a challenging task. Plagiarism detection tools are programs that compare document with possible sources in order to identify similarity and so discover submissions that might be plagiarized [4].

There are various tools available for plagiarism detection. These can be categorized on the type of text tools operates on:

- (1) Tools that check the source code
- (2) Tools that check the text.

Tools Available:

JPlag

JPlag is source code plagiarism detection tool started in 1997. Jplag is free but user must create an account. Jplag takes as input a set of programs, compares these programs pair wise (computing for each pair a total similarity value and a set of similarity regions), and provides as output a set of HTML pages that allow for exploring and understanding the similarities found in detail. JPlag [5] works by

converting each program into a stream of canonical tokens and then trying to cover one such token string by substrings taken from the other.

Moss

Moss stands for Measure of Software Similarity. It is a **system** developed in 1994 by Alex **Aiken**. Moss is source code plagiarism detection tool. Moss is also free but user must create an account it provides an internet service and have web interface. It produces both text and HTML reports.

JPlag and MOSS detect source code plagiarism based on textual analysis via the characteristic of source code. Both systems are freely available online. MOSS performs pair wise comparison via a fingerprint in each file, and finds the longest common sequence detected in the two fingerprints [6]. JPlag also finds the longest common sequence, but it first transforms the source code text. JPlag is available as a web service. JPlag has a powerful user interface for understanding the results. It has meanwhile inspired a similar interface for MOSS. JPlag is resource-efficient and scales to large submissions. MOSS is even better in this respect.

EVE2 (Essay Verification Engine)

It runs on Windows 2000, NT and XP systems and accepts text in several formats including: plain text, Microsoft Word, and Word Perfect. It produces a full report on each paper that contained plagiarism, including the percent of the essay plagiarized, and an annotated copy of the paper showing all plagiarism highlighted in red [7]. It is inexpensive to buy. It is designed to determine how much a document matches a single online source. It takes 2 to 45 minutes to scan a 5-7 page document. It performs adequately for short documents. According to Sebastian Niezgodna and Thomas P. Way, EVE2 detected 65% of Plagiarism for sample of 10 papers with high degree of plagiarism. It is downloaded on the system [8].

Plagiarism-Finder

This application compares the given document with sources on the Internet and generates HTML reports highlighting concurrent passages and providing links to the source, for verification. It runs on Windows 2000 and XP systems and accepts files in several standard formats such as PDF, DOC, HTML, TXT and RTF. At the time of writing (July 2009) a trial version is available [9] free, otherwise the price is \$125.

Ithenticate

The application compares a given document against the document sources available on the World Wide Web. It also compares the given document against proprietary databases of published works (including ABI/Inform, Periodical Abstracts, Business Dateline), as well as numerous electronic books and produces originality reports [10]. The originality reports provide the amounts of materials copied (in percentages) to determine the extent of plagiarism. No installation on home computer is required.

PlagiarismDetect

This is a freely available Internet service [11]. Users need to register by providing their names and email addresses. Once registered, text can be entered in the text box provided or a file uploaded for analysis. A report is then sent back to the user with a list of the links where the information has been copied from with percentages referring to the amounts copied.

Ephorus

This tool requires registration with the Ephorus site and, therefore, no downloads or installation is needed. Documents are submitted to the Ephorus website (www.ephorus.com). The search engine compares the given document to millions of others on the WWW and reports back with an originality report. License need to be purchased but the system can be freely tried [12]. It is widely used in Europe, South America and the U.S. by universities, colleges and secondary education.

Turnitin

It is internet based plagiarism detection software. Package is available for Windows 2000, NT and XP systems as well as for Mac OS X/9. The turnaround time for evaluation of report is several hours, although a fast service option is also provided. It provides three modules. OriginalityCheck shows how much percentage of paper matches the turnitin repository. GradeMark gives feedback by enabling editorial highlights, custom comments and editing marks on paper. PeerMark enables to learn from peers by their reviews [13]. Turnitin does not differentiate between correctly cited references and unacknowledged copying. Karl O. Jones Stated “It must be made clear that Turnitin should not really be considered a plagiarism detection system, it is merely a text matching system.”[14].

Summary

In this paper plagiarism detection tools are discussed. Plagiarism prevention methods without any doubt are the most significant means to fight against plagiarism, but implementation of these methods is a challenge for society as a whole. Education institutions need to focus on plagiarism detection methods.

There are various tools developed for plagiarism detection. But even the best detection tool can't detect better than human eye. No doubt, available tools help very much to detect plagiarism.

References

- [1] Stokes, F. and Newstead, S. (1995): Undergraduate cheating: who does what and why? *Studies in Higher Education* 20(2):159-172
- [2] .Culwin, F., MacLeod, A. and Lancaster, T., (2001): Source Code Plagiarism in UK HE Schools - Issues, Attitudes and Tools, Technical Report SBU-CISM-01- 01, South Bank University, 2001.
- [3] IEEE TV (2010). The IEEE Plagiarism Guidelines. Retrieved April 25, 2010, from IEEE: <http://www.ieee.org/portal/ieeetv>
- [4] Lancaster, T., F. Culwin. Classifications of Plagiarism Detection Engines. *ITALICS* Vol. 4 (2), 2005.
- [5] Lutz Prechelt, Guido Malpohl and Michael Philippsen (2000). JPlag: Finding Plagiarisms among a Set of Programs. <http://page.mi.fuberlin.de/prechelt/Biblio/jplagTR.pdf>.
- [6] Chang-Keon Ryu, Hyong-Jun Kim and Hwan-Gue Cho A Detecting and Tracing Algorithm for Unauthorized Internet-News Plagiarism Using Spatio-Temporal Document Evolution Model SAC '09 Proceedings of the 2009 ACM symposium on Applied Computing
- [7] Essay verification engine, EVE Plagiarism Detection System, WSEAS Transactions on Information Science and Applications Zaigham Mahmood. ISSN: 1790-0832 1357 Issue 8, Volume 6, August 2009
Available at: www.canexus.com/eve/index.shtml

- [8] Niezgoda S. and Way Thomas: SNITCH: A Software Tool for Detecting Cut and Paste Plagiarism. SIGCSE '06 Proceedings of the 37th SIGCSE technical symposium on Computer science education.
- [9] Plagiarism Finder
Available at:www.m4-software.de/en-index.htm
- [10] iThenticate,
Available at:www.ithenticate.com/static/home.html
- [11] PlagiarismDetect,
Available at:www.plagiarismdetect.com/
- [12] Ephorus,
Available at:www.ephorus.de/home_en.html
- [13] TurnItIn,
Available at: www.turnitin.com/static/index.html.
- [14] Jones K.: Practical Issues For Academics Using the Turnitn Plagiarism Detection Software. CompSysTech '08 Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing 2008
- [15] www.plagiarism.org
- [16] www.stateuniversity.com/blog/permalink/How-to-Avoid-Plagiarism.html
- [17] <http://en.wikipedia.org/wiki/Plagiarism>