ISSN:0975-9646

Survey on Different Plagiarism Detection Tools and Software's

Amruta Patil[1], Nikhil Bomanwar[2]
[1,2]*Computer Department, Mumbai University*

***Abstract*— Plagiarism happen in academics, paper publication, music, artwork growing rapidly, So the detecting plagiarism is very important. While the last few years plagiarism detection tools have been used mainly in research environments, sophisticated plagiarism software and tools are now rapidly emerging. In this paper, we provide an overview of different plagiarism software and tools to solve the plagiarism problem. We propose a feature classification scheme that can be used to study plagiarism detection software and plagiarism detection tools. This scheme is based on the software's general characteristics, tools characteristics, and tools attribute. We then apply our feature classification scheme to investigate 5 plagiarism software and 10 plagiarism detection tools which are either free or commercially available. Finally, we specify features that we consider important for plagiarism detection software and tools to possess in order to accommodate its users effectively, as well as issues that are either not addressed or insufficiently solved yet.**

***Keywords*— Put your keywords here, keywords are separated by comma.**

I. INTRODUCTION

1.1 OVWERVIEW and MOTIVATION

There are many types of plagiarism, such as copy and paste, redrafting or paraphrasing of the text, plagiarism of idea, and plagiarism through translation from one language to another. These types have made plagiarism one of the serious problems in academic area precisely. A modern research found that 70% of students confess to a few plagiarisms, with about half being guilty of an earnest cheating offence on a written assignment. Additionally, 40% of students confess to using the "cut- paste" method when completing their assignments. Differentiating between the plagiarized documents and non-plagiarized documents in an effective and efficient way is one main issue in plagiarism detection field.

Plagiarism can be found in the different areas such as literature, music, software, scientific articles, research papers, newspapers, advertisements, websites etc. A study carried in United States shows that among 18000 university students almost 40% of them have plagiarized at least once. According to Carroll [1], at least 10% of student's work is likely to be plagiarized in USA, Australia and UK universities [2]. Current methods of plagiarism detection are based on the characters matching, n-gram, chunks or terms.

As the use of internet increases plagiarism becomes challenge in school, university to maintain the academic integrity. So the use of efficient plagiarism detection tools has become very important in many higher education institutions, but the effectiveness of detection level depends on the type of algorithm and the type of obfuscation strategy employed by the plagiarist in order to create the plagiarized text.

The rest of this paper is organized as follows. Section 1.2 lists categories of plagiarism software that are intentionally included in this report. Section 1.3 lists some other survey tools which are included and also not included in this paper projects. Section 2 discusses the process of plagiarism. It describes the various the techniques for solving the plagiarism. Section 3 presents the review of existing plagiarism software and tools using this scheme. Section 4 draws some conclusions about the current state of existing discovery tools, and identifies some desirable features and characteristics that make detection tool truly useful, thus providing directions for future research.

1.2 SOFTWARE INCLUDED in THIS PAPER

1. PlagAware[7]
2. Plag Scan[8]
3. Ithenticate[10]
4. Check for plagiarism.net[9]
5. Plagiarism detection.org[11]

Table 1.2.1: Comparison of the mentioned five different software's based on their different features.

Feature	Plag Aware	Ithenticate	Plag Scan	Check for plagiarism.net	Plagiarism detection.org
Database Checking (online and offline)	Excellent	Excellent	Excellent	Very good	Very good
Internet Checking	Excellent	Excellent	Excellent	Excellent	Excellent
Publication Checking	Excellent	Excellent	Excellent	Good	Good
Multiple document comparison	Excellent	Excellent	Excellent	Very good	Very Good
Multiple languages support	Excellent	Excellent	Excellent	Excellent	Excellent
Sentence structure and synonym checking	Very good	Very good	Acceptable	Excellent	Acceptable

1.3 SOFTWARE TOOLS INCLUDED in THIS PAPER

1. Turnitin
2. EVE2
3. CopyCatchGold
4. Word Check[12]
5. Glatt
6. Moss
7. J Plag[5]
8. Google, Yahoo, Alta/Vista

Table 1.3.1: Comparison of different tools based on different attributes.

	Tools							
Attribute	Turnitin	EVE2	CopyCatchGold	Word Check	Glatt	Moss	Jplag	Google,Yahoo,Alta/ Vista
Check source code?	_	-	-	-	-	Y	Y	-
Checks free text?	Y	Y	Y	Y	Y	-	-	Y
Operates intra-corpaly	Y	-	Y	Y	-	Y	Y	-
Operates extra-corpaly	Y	Y	-	-	-	-	-	Y
Designed for use by student	Y	-	-	-	-	-	-	Y
Designed for use by teachers	Y	Y	Y	Y	Y	Y	Y	Y
Instant response	-	Y	-	Y	-	-	-	Y
Free	-	-	-	-	-	-	Y	Y
Accurate result	Good	Acceptable	Excellent	Unsatisfactory				
Operating environment	Web based	PC only	PC only	PC only	PC only			

Table 1.3.2: different software's that are not included in this paper and also the reason why they are no included

Tool	Included	Reason not included
Code Match	not yet	commercial (but free evaluation license possible)
P Detect	No	does not support Java
TEAMHANDIN	No	does not support Java
X plag	No	meant for inter-lingual plagiarism detection
yap3	No	does not support Java

II PLAGIARISM PROCESS

This section provides an introduction into the area of plagiarism, and background explanation for our feature classification scheme and the software feature tables in section 3.

2.1 PLAGIARISM PROCESS:

The process of plagiarism is very simple as copy the work of another or use the existing content.

The whole task of plagiarism includes

1. First search for the content from various resources that is text, pdf, web or any other resource.
2. Then use this content by copy and paste into the new document.
3. *After the new document we can say as plagiarized document.*

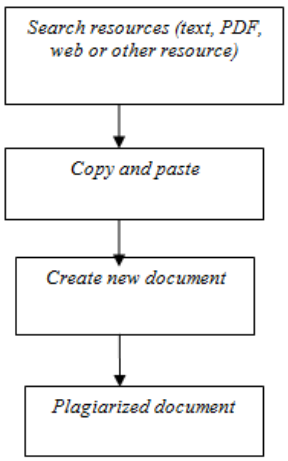

Fig.1 Basic steps for the plagiarism

Plagiarism detection process stages:

Lancaster and Culwin [6] define the important stages used for plagiarism detection as collection, analysis, confirmation and investigation. These four stages are important for designing error free process.

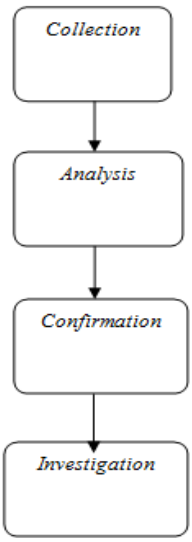

Fig. 2 Four-stage Plagiarism Detection Process

1. Collection

   This is the first stage of Plagiarism Detection Process, and it entails the student or researcher to upload their assignments or works to the web engine, the web engine acts as an interface between the students and the system.

2. Analysis

   In this stage all the submitted assignments are run through a similarity engine to determine which documents are similar to other documents. There are two types of similarity engines, first intra-corpal engine and second extra-corpal engine. The intra-corpal engines work by returning ordered list between each similar pairs. By contrast, the extra-corpal engines return suitable web links.

3.Confirmation

   The function of this stage is to determine if the relevant text has been plagiarized from other texts or to determine if there is a high degree of similarity between a source document and any other document.

4 Investigation

   This is the final stage of a Plagiarism Detection Process and it relies on human intervention. In this step a human expert is responsible for determine if the system ran correctly as well as determining if a result has been truly plagiarized or simply cited.

2.3 THREE MAIN TASKS OF THE PLAGIARISM DETECTION:

The detection process is divided into 3 tasks

1. Pre Processing
2. Intermediate Processing
3. Post Processing

Pre Processing:
It involves uploading source document and retrieving plagiarized documents from Corpus based on the source document. Once we acquire the specific data we send this data for intermediate processing.

Intermediate processing stage:
Involve the detection and comparison of the source and the plagiarized documents based on the algorithm.

Post Processing:
This is the final stage includes display the result that is document is plagiarized or not

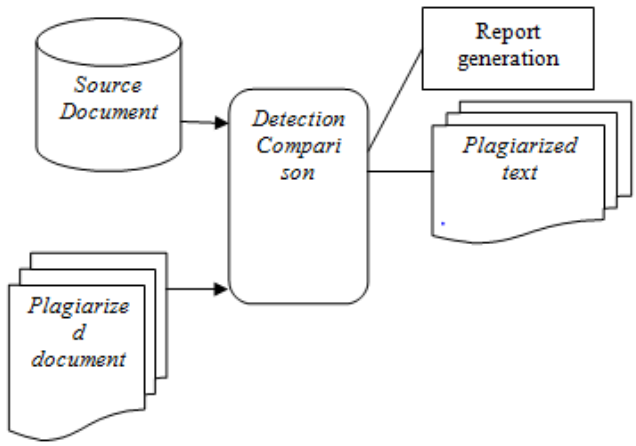

Fig .3 Generic retrieval process.

2.3 COMPARISION of DIFFERENT ALGORITHMS

There are some string matching algorithms which are used for detecting weather the document is plagiarized or not. Below tables shows the algorithms which are widely accepted and the difference between most efficient algorithm based on the time required for preprocessing, worst case complexity, execution time and accuracy of detection [3][4].

Table 2.3.1 Comparison of algorithms based on their accuracy

Algorithm	Preprocessing	Worst case	Execution time	Accuracy
Naïve algorithm	none	O(mn)	O(mn)	66.7%
Knuth Morris pratt	O(m)	O(n)	O(m+n)	65%
Boyre Moore	O(m+n)	O(mn)	O(mn)	75%
Rabin karp	O(n)	O(mn)	O(mn)	70%

III INVESTIGATION OF DIFFERENT PLAGIARISM SOFTWARE AND DIFFERENT DISCOVERY TOOLS

Table 1.2.1 shows the different plagiarism software for detecting plagiarism and characteristics of different software that is which features are check for plagiarism detection. We rank them starting from best that is plag aware we ranked them as follows, starting from the best, PlagAware, iThenticate, PlagScan,CheckForPlagiarism.net and lastly PlagiarismDetection.org. Table 1.3.1 shows detail attributes of plagiarism detection tools. All tools in table are grouped into specific tools, which were specially developed to detect plagiarism in submissions, and Internet search engines – alternative tools to detect suspected plagiarism. It is worth to point out that alternative tools haven't appropriate set of instruments to analyze suspected submissions qualitatively that is why these tools can't be viewed as serious plagiarism detection tools .

IV CONCLUSION AND FUTURE RESEARCH

In this paper we study the detecting plagiarism is very important not only in academics but also in industry, music, artwork etc. In particular, it has been shown in this study how the problem of plagiarism can be handled by using different techniques and tools. In this paper we saw that various software and tools are available for detecting plagiarism The comparison of the software and tools shown that still now their no software and tools that can detect or to prove that the document has been plagiarize 100%, because each software and tool has advantages and limitation, according to the features and performance described in the table. However there limitations in this software, tools which will affect the success of plagiarism detection significantly. The future work involves adding more capability and features to the current software and tools to detect the plagiarized document very efficiently.

REFERENCES

[1] J. J. G. Adeva, et al., "Applying plagiarism detection to engineering education," 2006, pp. 722-731.
[2] C. Lyon, et al., "Plagiarism is easy, but also easy to detect," Plagiary: Cross-Disciplinary Studies in Plagiarism, Fabrication, and Falsification, vol. 1, 2006
[3] Akhtar Rasool,Amrita Tiwari, Gunjan Singla,Nilay Khare "String Matching Methodologies:A Comparative Analysis" et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3394 – 3397
[4] Zeeshan Ahmed Khan1, R.K Pateriya2 "Multiple Pattern String Matching Methodologies: A Comparative Analysis" International Journal of Scientific and Research Publications, Volume 2, Issue 7, July 2012 1 ISSN 2250-3153
[5] L. Prechelt, G. Malpohl, and M. Phlippsen, "Jplag: Finding plagiarisms among a set of programs," Tech. Rep., 2000.
[6] F. Culwin and T. Lancaster, "Plagiarism issues for higher education," Vine, vol. 31, pp. 36-41, 2001.
[7] http://www.plagaware.com/about plagaware
[8] http://www.plagscan.com.
[9] http://www.checkforplagiarism.net
[10] http://www.ithenticate.com/index.html
[11] http://www.plagiarismdetection.org
[12] http://www.WordCHECKsystems.com