RESEARCH PAPER | Computer Science | Volume : 3 | Issue : 3 | March 2013 | ISSN - 2249-555X

Identification of Digital Copyright Violation: An Overview

KEYWORDS: Identification of digital Copyright violation, Plagiarism, plagiarism prevention, plagiarism detection, similarity measures

Mr. Deven J. Patel
Assistant Professor, Department of MCA Atmiya Institute of Technology and Science, Rajkot (Gujarat), India

Dr. Bankim L. Radadiya
Director of Information Technology, Navsari Agricultural University, Navsari (Gujarat), India

ABSTRACT *"Intellectual property theft" in the sense of Identification of digital Copyright violation and research work around the human art was produced. However, web, large databases, and is generally easy access to telecommunication, publishers, researchers and academic institutions to continue plagiarism is a serious problem. In this paper, we literally plagiarism (such as music, pictures, images, maps, technical drawings, etc., materials, plagiarism v) to focus on. We discuss the complex formal settings, then plagiarism detection software to report on some results and finally, the fact that plagiarism check-up rather than any serious unexpected side effects turn out. We believe that this paper all researchers, educators, and students to value and seminal work, which hopefully will encourage many still deeper investigation should be regarded as.*

INTRODUCTION

Intellectual property theft as "plagiarism" or "digital Copyright violation" has been around since the work produced by human art and research." Someone else's work as your own without plagiarism reference source can be defined as turning. In practice, different methods are commonly plagiarism. Some of them [13] to include:

- copy paste (copy word for word as literally) plagiarism
- paraphrasing (the same material in different words restating)
- translation plagiarism (translation used without reference to the content of the original work)
- artistic plagiarism (using the work presented in different media: text, images, etc.)
- the idea (using the same ideas that are not common knowledge) plagiarism
- Code plagiarism (reference without permission or using program code)
- a proper use of quotation marks (Failure to identify specific parts of the borrowed material)
- references incorrect information (wrong or nonexisting source by adding reference).

It is difficult to disagree about plagiarism problem of plagiarism is becoming more and more real society continued to increase knowledge and society's attention to the serious problem of piracy starts. More and more people realize that plagiarism amoral phenomenon that can not exist in a society with high ethical standards starts.

But why it's plagiarism problem becomes especially true only today? Although we make information technology age that makes our lives easier, but a set of problems all the time. The availability of digital documents (for example, easy access to the web) and plagiarism enrichment process is very simple and privately normally open good opportunities for turning in telecommunications fraud. [13] stated that, "Now that plagiarism is a serious problem for publishers, researchers and teachers to continue." The rest of this paper is as follows. Gives some ideas about plagiarism in the next section. Various methods are described for the plagiarism detection. After data analysis tools that are already developed. Finally, some conclusions are given.

WAYS HOW TO REDUCE PLAGIARISM

Nowadays, many methods developed in the fight against piracy and used. These methods can be divided into two categories. (1) methods for plagiarism detection

(2) methods for plagiarism prevention.

If we consider plagiarism as a kind of social illness then we can say that methods of the first class are precautionary measures which aim are to preclude rise of illness, but methods of the second class are cures which are aimed to avert existing illness. Some examples of methods in each class are as follows: plagiarism prevention - or punish honesty policies and systems, and plagiarism detection - automatically displays the plagiarism software tools. Each method has its application in determining the characteristics of a set. There are two main features that are common to all methods (see Table 1):

1) Work - intensity of method's implementation;
2) Duration of method's efficiency.

Work - the method of implementation of the magnitude of resources (especially time) to develop and use this method to bring in so far as it is necessary. Plagiarism prevention methods are usually time - consuming to recover, while plagiarism detection methods require less time.

Methods	Attribute of Methods	
	Implementation work – intensity	Duration of positive effect
Plagiarism prevention methods	Require more time to implement	Positive effect isn't momentary, but it is long term
Plagiarism detection methods	Require less time to implement	Positive effect is momentary, but it is short term

Table 1: Attributes of plagiarism detection and prevention methods

Duration of method's efficiency is that the length of time understanding the mechanism of the positive effect exists. Prevention methods to implement a long-term positive effect. In contrast, the invention provides methods of implementing short - term positive affects. Antipodal approach to a variety of methods to positively influence because of the methods used to fight against plagiarism - intimidation of society based on search methods, while plagiarism prevention methods against the tendency of society depends on change.

The fight against plagiarism - the difference between prevention and detection methods, although this is a common goal for all methods used. The fight for efficient, system plagiarism problem solving approach is required, which means that it is necessary to combine plagiarism prevention and detection methods. To achieve transient to - term positive re-

Volume : 3 | Issue : 3 | March 2013 | ISSN - 2249-555X

sults in plagiarism detection methods must be applied early stages of the problem, but no positive results achieved - time plagiarism prevention methods must be implemented. Plagiarism detection methods can not only reduce plagiarism, but plagiarism prevention methods phenomena can remove completely or at least reduced to a great extent. That is why plagiarism prevention methods without doubt the most significant measures to fight against plagiarism. That is why only a plagiarism detection methods and tools are discussed in this paper.

PLAGIARISM DETECTION METHODS

Plagiarism detection is usually based on a comparison of two or more documents. In order to compare two or more documents and information about the degree of similarity between them, it is a statistical value, therefore, is necessary to assign a similarity score for each document. These properties can be based on different metrics. There are many dimensions and aspects of the document, which can be used as a matrix. In this paper, we focus on a specific source code plagiarism detection metrics are used to not paying for, HALSTEAD matrix [6] like. The metrics used in most general purpose described in this section. Lancaster and Culwin their work [12] is used to search for plagiarism have attempted to classify metrics. They proposed two ways of how to classify metrics. The first classification matrix and the second one in the calculation process employing methods such as similarity search based on computational complexity is based on the number of documents involved. The classification matrix in the singular or pair or multi-dimensional matrix-matrix and corpal, depending on how many documents are in front, and a set of documents depending on the procedure involved, respectively, can be classified as. Operates on the entire corpus of documents to a corpal metric. Multidimensional metric operates on a number of selected documents. The classification matrix can be classified as superficial matrix and structural matrix. A square is a measure of the similarity metric that can be gauged simply by viewing one or more documents. This case does not require knowledge of natural language semantic features. A structural similarity metric that requires knowledge of one or more documents, it is a measure of the structure. Classification of the main theory of how the other metrics are built - they are, that is, 'semantically analysis or statistical methods based on the contents of the documents. Statistical methods to understand the meaning of a document requirements. As a general statistical approach document, words, compression matrix [8], Lancaster word pairs [11] and other metrics based on the values of the frequency description of the construction of vectors. Statistical metrics are language independent or [5] language sensitive may be. Purely statistical method of N-gram approach where the text of the N row [3] with a characteristic sequence of characters. Based on a statistical measure called the fingerprints of each document, where the n-gram is hashed and then try some fingerprints can be described with. Measures that are likely to be there. This action as a theoretical measure of [1], BM25 [15], the language and the model size. In many cases, the Euclidean distance between two document vectors between the similarity score is considered document. Documents equality is equal to zero. [9] The similarity of document vectors divided by their length can be calculated as the scalar product. The cosine of the angle between two document vectors seen from the equivalent of the original. In many cases, the document word frequency and word weight vectors that are made are automatically calculated for each document. Consider the word frequency [2] function is taken. Also cosine formula can have variation (see, Eq. 1), where also word weights are taken into account [2]:

$$S_{cos}(A,B) = \frac{\sum_{i=1}^{n}[\alpha_i^2 \times F_i(A) \times F_i(B)]}{\sqrt{\sum_{i=1}^{n}[\alpha_i^2 \times F_i^2(A)] \times \sum_{i=1}^{n}[\alpha_i^2 \times F_i^2(B)]}}$$

where αi – word weight vector; Fi(A), Fi(B) – frequency of the ith word in documents A and B, respectively.

Cosine function, proportion function, as well as dot production, Jaccard measure, Dice measure, overlap measure are symmetric similarity measures [2]. Symmetric or asymmetric similarity measures are one more classification. Asymmetric similarity measures are heavy frequency vector and heavy inclusion proportion model, which are derived from cosine function and proportion function by combining asymmetric similarity concept with heavy frequency vector [2]. Asymmetric similarity measures can be used for searching subset coping. Usually in different tools statistical methods are implemented due to their simplicity.

TOOLS FOR DETECTING PLAGIARISM

The authors concluded that "expanding the view of this problem, it is quite obvious that the teaching tools needed to automate and improve plagiarism detection." [12] According to the "plagiarism detection tools to compare programs that equality is possible to identify the source document and therefore it seems that submissions may be plagiarized."

There are tools available that can detect plagiarism in documents. The most popular plagiarism detection tools are Moss, JPlag, Turnitin, Eve2, CopyCatchGold, WordCheck, Glatt [4,10,12,13,14]. According to information available on the web leader between analytic search tools Turnitin is [7,8], due to the function. Each tool features that determine the application of a set. There are two main features that are common to all the tools:

1) Text tool operates on;
2) Type of corpus tool operates on

Tools can be divided into two groups according to the attribute "operates on the text tool type": means that the non-structured (free) tools that text and structured text (source code) are working on. In fact, free text or source code search tools are not limited to work on. It spreadsheets, diagrams, scientific experiments, music or any other non-verbal [12] corpora in the search for equality can be. Tools can be divided into three groups.

According to the attribute of "Corpus operates on the tool type": tools that work only intra corpus (where both the source and a copy of the documents within a corpus), tools that work only additional corpus (where a copy of the Corpus within and outside the source) and tools that both work - intra and extra corpus. [12]

Table 2 shows the characteristics of plagiarism detection tools in detail. Table, all the tools are tools, which were developed specifically to detect plagiarism in submissions and Internet search engine into - find alternate means of suspected plagiarism. It is worth believe that the optional equipment is suspected to play a proper set of submissions qualitative analysis of why these tools can not be viewed not as a serious plagiarism detection tool.

Attributes	Detection tools							
	Specific tools							Alternative tools
	Turnitin	Eve2	CopyCatchGold	WordCheck	Glatt	Moss	Jplag	Google Yahoo AltaVista
Type of text tool operates on								
Checks source code?	-	-	-	-	-	Y	Y	-
Checks free text?	Y	Y	Y	Y	Y	-	-	Y
Type of corpus tool operates on								
Operates intra-corpally?	Y	-	Y	Y	-	Y	Y	-
Operates extra-corpally?	Y	Y	-	-	-	-	-	Y
Other attributes								
Designed for use by students?	Y	-	-	-	-	-	-	Y
Designed for use by teachers?	Y	Y	Y	Y	Y	Y	Y	Y
Instant response?	-	Y	-	Y	-	-	-	Y
Free?	-	-	-	-	-	Y	Y	Y

Table 2: Attributes of plagiarism detection tools [based on 4]

Operation statistical or semantical methods of plagiarism detection tools, or both based on the idea of good results. Some search tools available descriptions can be concluded that the tools are a great part of the statistical methods to use to detect plagiarism, because these methods can also - they are easy to understand and implement the software.

Volume : 3 | Issue : 3 | March 2013 | ISSN - 2249-555X

It is emphasized that although the "plagiarism detection tools in detecting the matching text between documents provide excellent service, care needs to be taken in their use." plagiarism detection tools stolen from the text properly cited serious deficiencies in the text to distinguish the inability of these tools. That is why it is necessary for the human interposition of a paper before it is stolen by the public - and human judgment still manually checking [4] is required.

CONCLUSIONS

Plagiarism in the age of information technology has become more real and not a serious problem. The paper discusses how to reduce piracy. Education institutions need to focus on plagiarism detection methods. Widely used plagiarism detection methods are usually separate statistical analysis shows that the metrics are used because of their simplicity and easiness tools will be implemented. They have a number of shortcomings, and, therefore, still require manual inspection and human judgment. The human brain is a universal plagiarism detection tool, which semantically analysis of numerical(Alexandre, Duarte, & Marques, 2010) and verbal and non-verbal information for the document using methods that work with the. Capabilities of existing software tools for the detection of plagiarism are not available. According to "at least for now ... - nothing can completely replace human's watchful eye." But although a computer - based plagiarism detection tools can help you find documents significant for the theft.

REFERENCE [1] Aslam, J.A., M. Frost.(2003) An information-theoretic measure for document similarity. Proceedings of the 26th international ACM/SIGIR conference on research and development in information retrieval, pp. 449–450. | [2] Bao, J.P., J.Y. Shen, H.Y. Liu, X.D. Liu. (2006) A fast document copy detection model. Soft Computing - A Fusion of Foundations, Methodologies and Applications. Vol. 10 (1), pp. 41 – 46. | [3] Brin, S., J. Davis, H. Garcia Molina. (1995) Copy detection mechanisms for digital documents. ACM SIGMOD International Conference on Management of Data, San Jose, California, pp. 398–409. | [4] Delvin, M. (2002) Plagiarism detection software: how effective is it? Assessing Learning in Australian Universities, Available at: http://www.cshe.unimelb.edu.au/ assessinglearning/docs/PlagSoftware.pdf | [5] Gruner, S., S. Naven.(2005)Tool support for plagiarism detection in text documents. ACM Symposium on Applied Computing. pp. 776 – 781. | [6] Halstead, M. Elements of Software(1977) Science, Elsevier Publishers, New York. | [7] Schleimer, S., D.S. Wilkerson, A. Aiken.(2003) Winnowing: Local Algorithm for Document Fingerprinting. Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 76-85. | [8] Saxon, S.(2000) Comparison of plagiarism detection techniques applied to student code: Computer science project (Pt. II). Cambridge: Trinity College. | [9] Jones, E.W.(2001) Plagiarism monitoring and detection – towards and open discussion. In Proceedings of the twelfth annual CCSC South Central conference on The Journal of Computing in Small Colleges, pp. 229–236. | [10] Lancaster T., F. Culwin.(2000) A review of electronic services for plagiarism detection in student submissions. Paper presented at 8th Annual Conference on the Teaching of Computing, Edinburgh,. Available at: http://www.ics.heacademy.ac.uk/events/presentations/317_Culwin.pdf | [11] Lancaster T., F.(2004) Culwin. A visual argument for plagiarism detection using word pairs. Paper presented at Plagiarism: Prevention, Practice and Policy Conference | [12] Lancaster, T., F. Culwin. (2005) Classifications of Plagiarism Detection Engines. ITALICS Vol. 4 (2). | [13] Maurer, H., F. Kappe, B. Zaka.(2006) Plagiarism – A Survey. Journal of Universal Computer Sciences, vol. 12, no. 8, pp. 1050 – 1084. | [14] Neill, C.J., G. Shanmuganthan.(2004) A Web – enabled plagiarism detection tool. IT Professional, vol. 6, issue 5, pp. 19 – 23,. | [15] Robertson, S., S. Walker.(1994) Some simple effective approximations to the 2- Poisson model for probabilistic weighted retrieval. Proceedings of the 17th international ACM/SIGIR conference on research and development in information retrieval, pp. 232– 241.