Review of plagiarism detection and control & copyrights in India

Prof. S.S.Karadage
Department of Computer Science & Engineering/MSBTE /AIT Poly/ Vita/India
* *suryakantk105@gmail.com*

Prof. Mahesh S. Jadhav
Principal AIT Polytechnic/MSBTE /Vita/Maharashtra/India

Abstract

Plagiarism software's has been in use for almost a decade to get the sense of "theft of intellectual property". However, in today's digital world the easy access to the web, large databases, and telecommunication in general, has turned plagiarism into a serious problem for publishers, researchers and educational institutions. The first part of this paper discusses the different plagiarism detection techniques such as text based, citation based and shape based and compared with respect of their features and performance and an overview of different plagiarism detection methods used for text documents have taken. The second half of the paper is fully dedicated to copyrights in India.

Keywords — Plagiarism, copyrights, copy rights protection, plagiarism detection, citation, citation analysis.

Introduction

There are many definitions available for plagiarism, but the technical field uses the following one in large scale.

Plagiarism:

Plagiarism is derived from the Latin word "plagiaries" which means kidnapper. It is defined as "the passing off of another person's work as if it were one's own, by claiming credit for something that was actually done by someone else" [Wikipedia: Plagiarism]. Plagiarism is not always intentional or stealing some things from someone else; it can be unintentional or accidental and may comprise of self stealing. The broader categories of plagiarism include: There is a long list of plagiarism methods commonly in practice. [7] Some of these methodologies include

- **Copy-paste**: copying word to word textual contents.
- **Idea plagiarism**: using similar concept or opinion which is not common knowledge.
- **Paraphrasing**: changing grammar, similar meaning words, re-ordering sentences in original work. Or restating same contents in different words.
- **Artistic plagiarism**: presenting someone else's work using different media, such as text, images, voice or video.
- **Code plagiarism**: using program code, algorithms, classes, or functions without permission or reference.
- **Forgotten or expired links to resources**: addition of quotations or reference marks but failing to provide information or up-to-date links to sources.
- **No proper use of quotation marks**: failing to identify exact parts of borrowed contents.
- **Misinformation of references**: adding references to incorrect or non existing original sources.
- **Translated plagiarism**: cross language content translation and use without reference to original work.

Text based plagiarism

The text based plagiarism uses the vector space method; they use the fingerprints for each document for matching it with fingerprints in other documents and find out similarities. It also can calculate and count the redundancy of the word in the document. This method is suitable for non partial plagiarism as mentioned before use the whole document and use vector space to match between the documents, but if the document has been partially plagiarized it cannot achieve good results. It may include copy and paste, modification or changing some words of the original information from the internet book magazine, newspaper, research, journal, personal information or ideas [5].

A. Text based plagiarism detection process stages

1) Stage One Collection

2) Stage Two Analysis

3) Stage Three Confirmations

4) Stage Four Investigations [2]

B. Different methods used for textual plagiarism detection

1) *Grammar-based method:*

2) *External plagiarism detection method:*

Citation-based plagiarism

In academic and scholarly publications citations has been a valuable source of information about the content of the document. The Citation based plagiarism mainly focuses on the documentary similarities between the citations and references used in the research paper. The degree of similarity between citation patterns depends, among others factors, mainly on the amount of shared references (bibliographic coupling strength), and the extent to which the order of included citations, as well as their distance towards each other is similar.

Shape-based PD for flowchart

Most, if not all, discard the figures and charts before checking for plagiarism. Discarding the figures and charts results in look holes that people can take advantage. That means people can plagiarized figures and charts easily without the current plagiarism systems detecting it. There are very few papers which talks about flowcharts plagiarism detection. Therefore, there is a need to develop a system that will detect plagiarism in figures and charts. Flowcharts become a significant issue to explain different kinds of information based on figure types. In some documents, flowcharts are so important to illustrate a lot of details and make it easier to understand methodology of structured design is one of primary steps to build entire system and solving engineering problems that can be explained by using flowcharts and other types of figures. [6]

Comparison

The different plagiarism detection techniques such as text based, citation based and shape based are compared based on the two categories: first according to features and secondly according to performance Qualitative comparison used in comparing the features of software, where we are looking for properties of the methods used. Quantitative comparisons used in comparing the performance of the method used.

A. THE TEXT-BASED PDS

It delivers satisfying results if the plagiarized text is copied literally (copy paste), with minor alterations (e.g. Shake paste) or is machine translated. Text matching approaches continue to be suitable for detecting copy paste plagiarism, even for short passages. They are also advantageous in that they do not require citation information. The text based PDS, especially Ferret and W Copy find, which work with local document comparisons, deliver good results for identifying copy paste plagiarism given that the sources are available, however, if the text is paraphrased or translated by a human, the text based method yield a very poor performance. Thus they fail to identify e.g. paraphrased, translated and idea plagiarism.

B. CITATION-BASED PLAGIARISM DETECTION (CBPD)

It subsumes methods that use citations and references for determining similarities between documents in order to identify plagiarism. Compares the occurrences of citations in order to identify similarities. The most basic form is to measure the bibliographic coupling strength (citation overlap). Citation-based Plagiarism Detection is by no means a replacement for the currently used text-based approaches. CbPD must be carefully verified by humans, especially in cases where only a few citation Whereas the strength of existing PDS lies in detecting plagiarism on the sentence level in the form of identifying similar or identical consecutive words, the strength of the citation based approach lies in identifying translation- and idea-plagiarism or disguised paraphrasing Citations and citation patterns offer unique features that facilitate a PD analysis. They are a comparatively easy to acquire, language independent marker, since more or less well-defined standards for using them are established in the international scientific community. This information can be exploited to detect forms of plagiarism that cannot be detected with text-based approaches. [3]

C. SHAPE-BASED PD FOR FLOWCHART

This method detects flow chart figure plagiarism based on shape-based image processing and multimedia retrieval. The method managed to retrieve flowcharts with ranked similarity according to different matching sets. There are many plagiarism detection systems in which Most, if not all, discard the figures and charts before checking for plagiarism. Discarding the figures and charts results in look holes that people can take advantage. That means people can plagiarized figures and charts easily without the current plagiarism systems detecting it [6]. Therefore, there is a need to develop a system that will detect plagiarism in figures and charts there is a need to develop a system that will detect plagiarism in figures along with their contents. There is a need to develop a method for detecting figure plagiarism based on shape-based image processing where different types of figures can be considered for method for detecting figure plagiarism.

COPYRIGHTS IN INDIA

WHAT IS COPYRIGHT?

Copyright is a right given by the law to creators of literary, dramatic, musical and artistic works and producers of cinematograph films and sound recordings. In fact, it is a bundle of rights including, inter alia, rights of reproduction, communication to the public, adaptation and translation of the work. There could be slight variations in the composition of the rights depending on the work.

WHY SHOULD COPYRIGHT BE PROTECTED?

Copyright ensures certain minimum safeguards of the rights of authors over their creations, thereby protecting and rewarding creativity. Creativity being the keystone of progress, no civilized society can afford to ignore the basic requirement of encouraging the same. Economic and social development of a society is dependent on creativity. The protection provided by copyright to the efforts of writers, artists, designers, dramatists, musicians, architects and producers of sound recordings, cinematograph films and computer software, creates an atmosphere conducive to creativity, which induces them to create more and motivates others to create

THE CLASSES OF WORKS FOR WHICH COPYRIGHTS PROTECTION IS AVAILABLE IN INDIA

Copyright subsists throughout India in the following classes of works:
  o   Original literary, dramatic, musical and artistic works;
  o   Cinematograph films; and
  o   Sound recordings.

Artistic work

A painting, a sculpture, a drawing (including a diagram, map, chart or plan), an engraving or a photograph, whether or not any such work possesses artistic quality; a work of architecture; and Any other work of artistic craftsmanship.

Musical work "Musical work" means a work consisting of music and includes any graphical notation of such work but does not include any words or any action intended to be sung, spoken or performed with the music. A musical work need not be written down to enjoy copyright protection.

Sound recording "Sound recording" means a recording of sounds from which sounds may be produced regardless of the medium on which such recording is made or the method by which the sounds are produced. A phonogram and a CD-ROM are sound recordings.

Cinematograph film "Cinematograph film" means any work of visual recording on any medium produced through a process from which a moving image may be produced by any means and includes a sound recording accompanying such visual recording and "cinematograph" shall be construed as including any work produced by any process analogous to cinematography including video films.

Government work "Government work" means a work which is made or published by or under the direction or control of
  o   the government or any department of the government
  o   any legislature in India, and
  o   Any court, tribunal or other judicial authority in India.
\

Indian work

"Indian work" means a literary, dramatic or musical work,

- o the author of which is a citizen of India; or
- o which is first published in India; or
- o The author of which, in the case of an unpublished work is, at the time of the making of the work, a citizen of India.

AUTHORSHIP AND OWNERSHIP

Copyright protects the rights of authors, i.e., creators of intellectual property in the form of literary, musical, dramatic and artistic works and cinematograph films and sound recordings. Ordinarily the author is the first owner of copyright in a work.

Who is an author?

In the case of a literary or dramatic work the author, i.e., the person who creates the work.

- In the case of a musical work, the composer.
- In the case of a cinematograph film, the producer.
- In the case of a sound recording, the producer.
- In the case of a photograph, the photographer.
- In the case of a computer generated work, the person who causes the work to be created.

DIFFERENT RIGHTS

In the case of a literary work (except computer program me), copyright means the exclusive right

- To reproduce the work
- To issue copies of the work to the public
- To perform the work in public
- To communicate the work to the public.
- To make cinematograph film or sound recording in respect of the work
- To make any translation of the work
- To make any adaptation of the work.

RIGHT OF COMMUNICATION TO THE PUBLIC

Communication to the public means making any work available for being seen or heard or otherwise enjoyed by the public directly or by any means of display or diffusion. It is not necessary that any member of the public actually sees, hears or otherwise enjoys the work so made available. For example, a cable operator may transmit a cinematograph film, which no member of the public may see. Still it is a communication to the public. The fact that the work in question is accessible to the public is enough to say that the work is communicated to the public.

REGISTRATION OF COPYRIGHT

Acquisition of copyright is automatic and it does not require any formality. However, certificate of registration of copyright and the entries made therein serve as *prima facie* evidence in a court of law with reference to dispute relating to ownership of copyright.

THE PROCEDURE FOR REGISTRATION OF A WORK UNDER THE COPYRIGHT ACT, 1957

Copyright comes into existence as soon as a work is created and no formality is required to be completed for acquiring copyright. However, facilities exist for having the work registered in the Register of Copyrights maintained in the Copyright Office of the Department of Education. The entries made in the Register of Copyrights serve as *prima-facie* evidence in the court of law. The Copyright Office has been set up to provide registration facilities to all types of works and is headed by a Registrar of Copyrights and is located at B.2/W.3, C.R. Barracks, Kasturba Gandhi Marg, New Delhi- 110 003, Tel: 338 4387
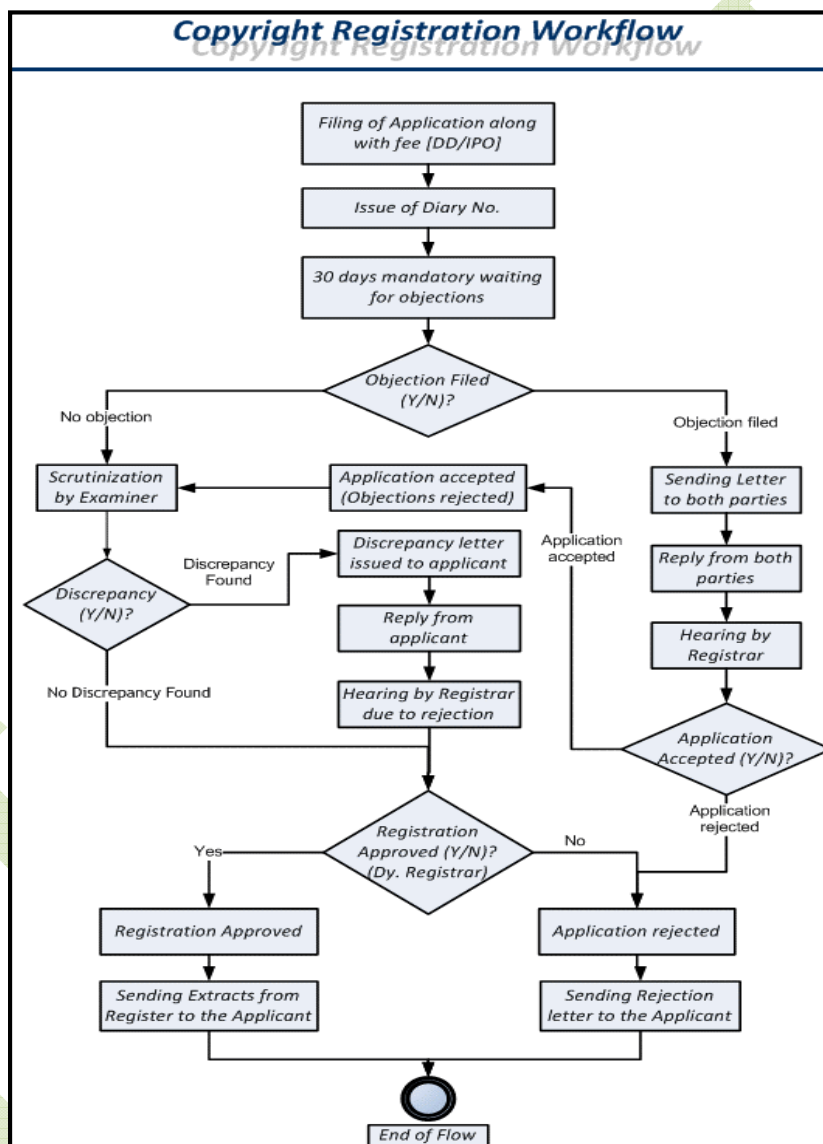

Figure. 1 Copyrights registration workflow [8]

Copyright infringements

The violations of copyright laws is a major issue today the literacy about copyrights and its laws should be done on a large scale by the governing bodies. The following are some of the commonly known acts involving infringement of copyright:

- Making infringing copies for sale or hire or selling or letting them for hire;
- Permitting any place for the performance of works in public where such performance constitutes infringement of copyright;
- Distributing infringing copies for the purpose of trade or to such an extent so as to affect prejudicially the interest of the owner of copyright ;
- Public exhibition of infringing copies by way of trade; and
- Importation of infringing copies into India.

Copyright piracy study

A study on copyright piracy in India has been carried out by Government of India which was sponsored by Ministry of Human Resource Development in 1999 the report is available on HRD ministry website the committee came up with some conclusions and suggestions which are enlisted below The study focused on the issues relating to the problem of copyright piracy in India. It attempts to arrive at a firsthand assessment of the piracy phenomenon and covered the main copyright segments namely cinematographic works (including video), sound recording, computer software, literacy works and the performers. Besides examining why and how piracy occurs, it tries to assess its economic impact in the country. The main objective is to suggest measures to effectively tackle this malady.

Recommendations

1. Since the direct loser due to copyright piracy are the right holders, the prime responsibility of protecting their copyrights lie with the right holders themselves. Firstly, the right holders should take enough precaution to protect copyright works. In case violations come to their notice/knowledge, they should file complaints with the police. They should also help the police in conducting raids and producing evidence (e.g. proof of ownership in works) during the trial by the court.

2. The copyright industry associations/copyright societies should launch an extensive campaign through print and electronic media highlighting the adversities associated with the piracy. Lectures, seminars, workshops etc. could be organized in schools, colleges, universities and other places to create a consciousness among people against the evils of piracy. The message should be conveyed in clear terms that in the long run piracy is against the interest of all in the society excepting the pirates.

3. The law enforcement authority like police needs to be imparted proper training in copyright fields. Apart from telling them how to differentiate original copyright products from the pirated ones, the various provisions of the Copyright Act are also to be taught. [8]

CONCLUSIONS

The paper explains three different methods used for plagiarism detection. The Text-based PDS convince in detecting local forms of plagiarism, such as short passages of copied or only slightly paraphrased text. In contrast, they fail, to detect paraphrased and translated plagiarism. The citation-based approach is based on citation analysis and allows duplicate and plagiarism detection even if a document has been paraphrased or translated. Shape based plagiarism for flowchart presents a method for detecting flow chart figure plagiarism based on shape-based image processing but fails to detect plagiarism between different types of figures. Thus plagiarism detection system should not be based on single method but must be based on the combination of different plagiarism detection methods. Also it takes a light on the copyrights in India, protections, rights, violations of laws.

References

[1] Hermann Maurer, Frank Kappe, Bilal Zaka" Plagiarism - A Survey" Institute for Information Systems and Computer Media Graz University of Technology, Austria, vol. 12, no. 8 2006

[2] Ahmed Hamza Osman, Naomie Salim1, and Albaraa Abuobieda," Survey of Text Plagiarism Detection" International University of Africa, Faculty of Computer Studies, Khartoum, Sudan Vol. 1, No. 1, June 2012

[3] Bela Gipp OvGU," Citation-based Plagiarism Detection – Idea, Implementation and Evaluation "Germany / UC Berkeley, California, USA 2010.

[4] Bela Gipp, Norman Meuschke, Joeran Beel" Comparative Evaluation of Text- and Citation-based Plagiarism Detection Approaches using GuttenPlag"

[5] Asim M. El Tahir Ali, Hussam M. Dahwa Abdulla," Overview and Comparison of Plagiarism Detection Tools "Department of Computer Science, Germany & UC Berkeley JCDL 2011.

[6] Senosy Arrish, Fadhil Noer Afif, Ahmadu Maidorawa and Naomie Salim" Shape-Based Plagiarism Detection for Flowchart Figures in Texts" Faculty of Computing, University Technology Malaysia, Skudai, Malaysia (IJCSIT) Vol 6, No 1, February 2014

[7] ccc.inaoep.mx/~villasen/bib/Plagiarism%20-%20A%20Survey.pdf

[8] http://copyright.gov.in/Documents/handbook.html