IPR, Plagiarism and the Text Data Security Pyramid

Rakesh Kumar Mishra†
School of Management Sciences, Varanasi
and
Rohit Ramesh
United Institute of Management, Naini, Allahabad

Received 11 July 2006, revised 18 August 2006

Plagiarism is a critical issue of concern in modern times and a major threat for evolving and preexisting text literature available in public domain. Adding fuel to the existing threat is the absence of any legal framework for preserving folklore. It being equipped with the capability of pattern matching and recognition supplemented by genetic algorithms can effectively be utilized for this purpose. It is the imperative need of society that these plagiaristic acts are restricted and the sanctity of existing work maintained. The paper discusses the predominant practices followed for data protection and modern methods of communication.

Keywords: Plagiarism, watermarking, pattern recognition and matching, genetic algorithm, folklore

Kavya Viswanathan, an American undergraduate student at the Harvard College and a novelist who was offered two book deal for $500,000 by the publisher Little Brown & Co, which Dream Works planned to develop into a film, caused furor when shortly after her first publication of novel ‘How Opel Mehta Got Kissed, Got Wild & Got a Life’, she was declared a plagiarist. Some of the extracts that raised alarm had a high order of similarity with another author, McCafferty’s work. Moreover, there were other portions which resembled extracts from ‘Haroun and the sea of stories’ by Salman Rushdie and ‘Can you keep a secret’ by Sophie Kinsella. Such instances reiterate the real threat of plagiarism and the extensive effort that is required to ensure the sanctity of original author’s work. Another concern is preexisting literature like folklore of India, which is rich not only in volume but also in values. It is essential to take care that extracts of ‘Ramayana’, ‘Shreemad Bhagwat Geeta’ and other epics which are upheld the Indian cultural heritage are not used by plagiarist and claimed as their own work.

Tracking an act of plagiarism is exceptionally labour intensive and requires rigorous comparison of the original and copied manuscript and track out content similarity vis-à-vis context similarity. The most viable options appear to be data hiding or data protection techniques, or a combination of both, applied to all publicly available documents, texts and scriptures for upholding the rights of ownership.

Data Hiding

Data hiding essentially means ‘to restrict the accessibility of data to such an extent that prima facie the existence of data itself is not visible’. Data hiding involves embedding data into digital media for the purpose of identification, annotation, and copyright. Important uses of data hiding in digital media include providing proof of copyright, and assurance of content integrity. This technique is prominent in practice whenever a company/person needs to protect any dataset that is vulnerable to tampering and may lead to communication of wrong information. It is thus a form of protection to ‘right of privacy’ in a more simplistic parlance.

One of the ways to use data hiding is to conceal a piece of information into a host dataset which can be used later on for the authentication of any sort of proprietary claims. This technique is identical to the ancient style of cryptic writing but differs in the approach because of the fact that the content hidden information is imperceptible to human vision. This approach is mandated under the Article 11 and 12 of the WIPO Copyright Treaty, that prescribes adequate legal protection for both, rights management system and technological anti-circumvention measures.
Data Protection

Data protection refers to ‘an act of ensuring intactness of data through any means’. Data with advent of computers and digitization has become a most vulnerable asset, sensitive to tampering. Data has been given the status of property (assets) under Section 22 of the Indian Penal Code, 1860 and thus entitled for protection under the same. Trade-Related Aspects of Intellectual Property Rights (TRIPS) Agreement also defines data and prescribes measures for its protection. In India, data property is fully protected under the Copyright Act of 1957 and IT Act 2000 (ref. 3).

The main objective of both the above principles is to maintain the sanctity of critical data elements which if exposed and tampered would lead to misinformation. With the expansion of technology from wired to wireless, the span of data accessibility is on an ever increasing curve and so is the potential of data tampering and manipulation. The other aspect of preventing miscommunication is to protect data from persisting threats that are increasing and upgrading exponentially, through a robust mechanism to trace infringements.

Need of ‘right of privacy’ however, needs to be balanced with ‘freedom of information’ which is in turn protected under Article 19(2) of IT Act 2000, which equates right to impart and receive information to that of freedom of expression and declares it as one of the fundamental rights of a citizen. This implies that not all kinds of data can be subjected to data hiding and protection.

Data Under Threat

With the advent of mass level computerization, globally, there has been a paradigm shift in the methodology of communication and warehousing of information, and technology has acquired the centre stage in the effort of real time communication. An evaluation of the modern communication method poses certain distinct challenges when it comes to information security and authenticity and needs careful examination.

Multimedia, one of the modern methods, refers to a computer-based interactive communication process that incorporates text, graphics, sound, animation and video i.e. a system that involves all the human senses towards communication. The interactivity and non-linearity of communication process are prime characteristics of any multimedia system. Multimedia is extensively used by industries for the purpose of communication.

Education and Research

Multimedia is utilized for effective teaching-learning process. Besides being interactive, it also, through its inherent characteristics, allows exclusive demonstration of the topic of interest. It is used extensively by research scholars to demonstrate their work in seminars and conferences, and exhibit the results in a self-explanatory manner.

Corporate Communication

Multimedia presentations are used for communication of strategies, targets and decisions besides the conventional text dispatches. This helps in better communication besides helping in quicker reviews and better participation among the decision makers and decision implementers for evolution, evaluation and implementation of strategy.

Modern day corporate institutions prominently use the digital electronic devices and media for communication and storage of data, due to their various inherent advantages, yet at the very same time there exists a distinct disadvantage in this system of communication, which is that the system is capable of producing indistinguishable number of copies of the digital data. This needs consideration due to certain characteristics like its quality of signals being higher in comparison to analog, low space requirement and adaptability towards error correction. As a result, digital content, copied as well as the original, produce same quality signal thus making comparisons very difficult. At the same time with ever-growing span and reliability of the digital public network, the Internet –locating the point of infringement is again a problem. If someone infringes the system and propagates wrong and/or altered information; by the time it would be noticed and taken care of, the damage to reputation, prestige and trade may already be done.

Hence, such modern means of data communication very much qualify for incorporation under laws in relation to data. Though there are individual acts catering to different aspects like the Copyright Act, Right of Privacy, IT Act 2000, etc., effective implementation pose a problem. Further, there are other procedural problems like how to proclaim that an author is a plagiarist or that the information in circulation is unauthentic or for that matter authentic. The other aspect is that of translation of Indian epics
into multimedia digital form and their global accessibility for social upliftment and enlightenment.

Globalization is having an alarming influence over the economic and cultural heritage of a country. India today, for example, is being noticed like never before, for commercial as well as non commercial reasons leading to an over-exposure of our cultural and social heritage world-over, thereby increasing the probability of an act of plagiarism with respect to our cultural and social heritage. It is an eye opening fact that there does not exist any Convention, Recommendation or any other text that fully addresses the need to safeguard the country’s priced folklore.\(^{10}\)

Protection of information available publicly i.e. released for compliance of Freedom of Information may also require special attention. Towards such Information, IT can be a very constructive partner. Ever since concerns for IPR and profitability has acquired great importance, the entire global intellectual community is working towards evolving a system for development and maintenance of IPR. The system does not necessarily mean only human resource, but also incorporates technology wherein, human resources with their diversified capacities act as multipoint distributed system each having specialization in respective field on the other hand the technology. Specifically IT will act as an interactive subordinate to enhance the capacity of conversion of knowledge into product.\(^{5,11}\)

**Data Security Measures**

IT does have potential to adapt and employ some of the domains that typically fit into the arena of security like:

**Cryptography**

It is the art of converting original information into a non-interpretable form so that, only the one with precise knowledge about the deciphering process and input sequence, can actually read it. It literally means ‘art of secret writing’. This technique does not exactly fit within the gamut of IPR regulation, which discourages circumventing of information using IT tools.\(^{8}\) The main problem again here is that since it is very difficult to distinguish between the copy and the original, proving infringement is an onerous task. Cryptography is prominently used for data transmission to avoid data tapping over the physical media and has been efficiently effective there too. The major threat is that a cryptic content immediately grabs the attention of the crypto analyst and thereby encouraging them to break the code, and with computers and host in pure digital form, this is not a cumbersome act at all.\(^{12}\)

**Steganography**

It is derived from Greek words ‘stegeno’ and ‘graphes’ which together means ‘undercover writing’. This is a variation of the cryptography but unlike cryptography, the received data does not appear to be alien and easily escapes the attention of cryptoanalysts. In steganography, dummy information encapsulates the actual information, which has to be extracted at the receivers end; e.g. an educational or patriotic movie, may contain text message for intelligence units; while the rest of the world is viewing the movie the intelligence agencies are exchanging vital information over public broadcast media. Here too, the actual content is circumvented by other content.\(^{12}\)

**Watermarking**

It is the art of carving hidden or visible signs within the content, say for example, currency notes. In India, there are several marks embedded in currency notes that are visible through naked eyes or using special light devices for viewing. These signs are used for authentication of the currency and restricting fake currency from intruding in the market. Another example is a marksheet wherein a special monogram is hidden, which is visible only when the photocopy of the marksheet is made, copy thus produced contains the visible image of hidden monogram and is self authenticated and does not require authentication. Watermarking does not come under data hiding as the data which is to be conveyed remains in actual form and is not circumvented by any means, rather is protected by the visible/invisible mark as a sign of authority.\(^{12}\) The main objective of embedding a watermark to any data set is to ensure the imperceptibility of the mark to such an extent that even digitally compared copies (unmarked original vis-à-vis marked original) can’t be distinguished remarkably.\(^{12}\)

**Conclusion**

Protection of current work (text) that belongs to an author is relevant but at the same time it is also essential to evolve strategies to protect preexisting literature and written traditions. For the enforcement of IPR it becomes imperative that there should exist a
sound system to identify and enforce protection against piracy. Here IT can play a significant role due to its tremendous potential.

References
1 James Rambaugh et al., Object Oriented Modelling and Design (Prentice-Hall Of India Private Limited, New Delhi), 1998.
7 Information is processed outcome of data, namely, the raw facts figures or statistics.