An Effective Methodology For Removing Pirated Content & to Identify Digital Copyright Infringement

KEYWORDS

Dr. Kishor H. Atkotiya	Mr. Nilesh Pratapsinh Gohil
Head, Department of Computer Science, J H Bhalodia Women's College, Rajkot - Gujarat (India)	Head & Assistant Professor, Department of Computer Science, Shri V. J. Modha College of IT, Porbandar - Gujarat (India)

ABSTRACT Cyber-pirates are the people who make and distribute copies of a music files, videos, video games, films, books and other digital content on internet illegally without authorization from the developers or owners. Content publisher such as film industries and publications are expecting to lose several billions of dollars in copyright revenues. All these things promoted by the search engines for earning revenues. To address this problem, I propose building an effective methodology, where content publisher will register their valuable digital content. This web service will crawl the web, than compare the web content to the registered content and will notify the content owners as well as search engines of illegal intellectual property.

Protecting intellectual property has received a lot of attention recently, both in terms of revised intellectual property laws, as well as new technology-based solutions. We now outline a few technology-based solutions in research proto-types, or that are already available in commercial products.

Fundamental

Plagiarism has been defined as "the taking and using as one's own of the thoughts, writings or inventions of another". There are many varieties and degrees of plagiarism. We will deal here with plagiarism of written work in academia and science, although the problem is not limited to these areas. Many people think of plagiarism as copying another's work, or borrowing someone else's original ideas. But terms like "copying" and "borrowing" can disguise the seriousness of the offense.

According to the Merriam-Webster Online Dictionary, to "plagiarize" means

- To steal and pass off (the ideas or words of another) as one's own
- To use (another's production) without crediting the source
- To commit literary theft
- To present as new and original an idea or product derived from an existing source.

In other words, plagiarism is an act of fraud. It involves both stealing someone else's work and lying about it afterward.

All of the following are measured plagiarism:

- Turning in someone else's work as your own
- Copying words or ideas from someone else without giving credit
- Failing to put a quotation in quotation marks
- Giving incorrect information about the source of a quotation
- Changing words but copying the sentence structure of a source without giving credit
- Copying so many words or ideas from a source that it makes up the majority of your work, whether you give credit or not (see our section on "fair use" rules)

Problems in Focus

Plagiarism, either intentional or inadvertent can occur in varying degrees for quite a few reasons:

- Some don't know they are or don't understand how to properly cite a source to give credit.
- Laziness-don't want to do the work and Poor planning-ran out of time.
- Writers block.
- Competition-they think others are and don't want them to have the advantage.
- Too busy to do the work.

For example, the web also creates new problems for publishers of digital content. Once a customer has purchased some goods and pay for them, the merchant to deliver the content. Digital case, the trader such as HTTP or email to use some of the information exchange protocol can deliver the goods to the customer. However, once a provider of digital content delivery, customers own Web site on this material, a Usenet newsgroup can post it or email it to friends or perhaps free or a reduced price. Publisher loses revenue due to reduced sales, although other potential customers of the cyber - rover will begin to access the content. Intellectual property protection has received much attention recently, both refined and intellectual property, as well as new solutions based on the laws in terms of technology. I am going to buildup research prototype of technology-based solutions that reduce plagiarism through removing content from the search engines. Figure 1.1 shows the reality of the search engine and plagiarism.
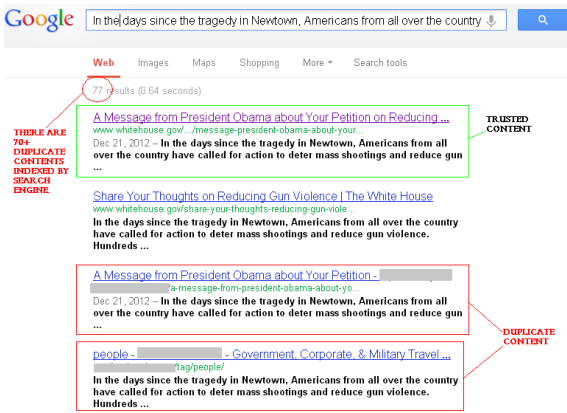
Figure 1.1 Problems in Focus

Recent approaches to combat cyber-piracy
- Plagiarism Prevention
- Plagiarism Detection

Plagiarism Prevention:
With copy prevention, content providers make it difficult to copy the original content. One common approach is to place information on stand-alone CD-ROM systems and allow users to access the data only through a restricted interface. A related approach is to use special purpose hardware to build fully trusted systems using fully trusted components and secure interconnections.

We believe that protecting content using special-purpose hardware works well in practice for relatively small-scale deployments. However this approach is unlikely to work for a global digital distribution channel because of the following "chicken and egg" problem. End-users will buy special-purpose hardware units only if (1) there is significant, valuable content available through these units, and (2) this content is not available elsewhere (e.g., as physical PDS). Publishers on the other hand are unlikely to switch away from existing distribution media (e.g., physical PDS), until they are convinced that customers will quickly adopt the new means (recall failure of Sony's Betamax video players and Digital Video Express' DivX players). Firstly, their approach requires a user's favorite viewer (Like emacs, vi, Microsoft Word) to be DigiBox-compatible. Secondly, hackers with software emulators or screen capture software can copy the data when they view or play it. Also, the approach is designed for a publisher who is exclusively selling content online.

Plagiarism Detection:
With copy detection, the content provider (author or publisher) imposes few restrictions on content distribution unlike the copy protection schemes we discussed earlier.
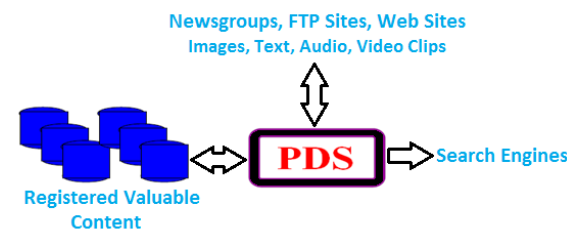

Figure 1.2 Plagiarism Detection Systems

However the provider registers his/her valuable digital content into a Plagiarism Detection System (PDS). The PDS then accesses public web sites, newsgroups, and FTP sites, and notifies the publisher as well as SEs if it finds copies of registered digital content using one of the following techniques:

Watermarking: In the watermarking approach, the content provider uses Steganographic techniques to hide information into digital content before selling the content. For instance when the content provider sells an image to a customer, the image is imperceptibly marked with the user's credit card number as well as the content publisher's name. When the PDS finds digital content at a public site, it automatically extracts the watermarks from the content. The PDS then notifies the corresponding publisher where the copy is located as well as the extracted credit card number (i.e., who was the initial buyer of the content.) Currently, there is a significant body of research into watermarking for multimedia content. For example, if two customers collude to compare the different versions of the content they bought, they can automatically remove differences (the watermarks) between the two versions. Also the watermarking approach does not address the problem of a person digitizing content available through other means (e.g., ripping a CD), and making it publicly avail.

Content-based: In this approach, the PDS compares the content on public sources (e.g., web sites) against the registered content and identifies potential copies of the registered content. For example, if Walt Disney Corporation wants to find all images of Mickey Mouse available on the web, they would register some sample images of Mickey Mouse with the PDS. The PDS will then try to identify images on public sources that are "similar" to Mickey Mouse and notify Disney of these copies. Building a content-based PDS is easy if the PDS only has to find exact copies of registered content. For example, we can use a simple checksum-based comparison of the content to identify potentially exact copies and then manually compare the content.

However, a cyber-pirate can defeat such a PDS by modifying his copy of the content so it has a different checksum than the original registered content. For instance, the cyber-pirate may change the sampling rate of an audio clip before offering it on a web site. In general, the cyber-pirate will modify digital content so that the "quality" of the modified content is retained and there is little loss in the content's commercial value. Therefore, the goal of the PDS is to compare digital objects efficiently and identify good-quality renderings of registered content.

One problem with the copy detection approach is that the PDS can only access digital content in public web sites and newsgroups. The PDS cannot find copied content that is delivered through private email or content that is available behind fire-walls and password-protected sites. However, we believe this weakness is not significant for the following reason. Content publishers expect to lose significant revenues primarily from piracy on publicly accessible sites and not from "below the radar" password-protected sites and firewalls. If a password-protected site tries to attract more traffic by publishing its password, the PDS can also access content at the site. So we expect PDS systems to play a crucial role in protecting intellectual property as we discuss in the next section.

CONCLUSIONS

Digital Copyright Infringement in the time of data innovation has gotten to be all the more real and not a difficult issue. The paper talks about the most effective method to decrease digital copyright infringement. Training foundations need to concentrate more on digital copyright infringement identification methods. Generally 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. Abilities of existing tools for the detection of copyright infringement are not accessible. But even though a computer - based plagiarism detection tools can be of assistance for you to locate documents significant for the pilfering.

REFERENCE [1] 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. | [2] 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. | [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/ assessing learning / 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,. |