

Submitted by Morris Rosenthal

## **Introduction**

The Green Paper rightfully acknowledges that the Internet presents copyright challenges and opportunities that require a rethinking of current copyright law. It falls short, however, in addressing the failure of the Digital Millennium Copyright Act to offer practical protection of copyright for the individual authors, photographers, musicians and video producers who are lauded as the lifeblood of our nation's creativity.

The hazard of Internet infringement is not the act of making copies. Whether an individual, or a billion individuals for that matter, make copies of a creator's work isn't the real issue. Individuals have been copying books by hand, replicating works of art, using reproduction technology from reel-to-reel tapes to digital memory to copy music and video for as long as those options have existed. What makes the Internet unique is not the ease of copying and distributing work, it's the ease of discovering these illicit copies, which short-circuits the legitimate distribution that finances the entire creative endeavor. Even when kids taped songs on cassette off the radio, the creators received some payment from the radio stations. Today, the creators are cut completely out of the loop.

The Green Paper speaks to the idea that discovery on the Internet opens doors and opportunities for creators, but it ignores that the same discovery process often negates the opportunity for creators to earn a living from their work. As an author who has wasted hundreds of hours fighting against Internet infringements by way telephone, fax, e-mail, and filing thousands of URL complaints through Google's DMCA Dashboard, I can attest that it's a whack-a-mole game that only ends when the source material (the original Internet page copyrighted by the author) has lost its own discovery status and no longer attracts thieves.

Google is the closest thing to a monopoly in the business of Internet discovery, yet even today, Google will refuse to act on DMCA complaints against infringements in their search engine indexed on larger websites, stating:

"We have received and reviewed your complaint. At this time, Google has decided not to take action based on our policies concerning content removal. As always, we encourage you to resolve any disputes directly with the owner of the website in question."

Google's policy is in violation of the DMCA which does not give search engine operators the leeway to decide whether or not they feel like removing infringing content from their search results, yet they take the position that they will not do so unless compelled by a court of law. For an individual author to have to pursue a Federal Court case against one of the world's largest and best funded companies is a highly stressful and expensive undertaking, to say the least. No explanation of a refusal is ever forthcoming from Google, they may acknowledge and act on a follow-up or they may not.

The remainder of this document is an abbreviated recap of feedback I filed with the copyright office for their solicitation related to establishing a small claims court for copyright complaints.

## **Why Web Page Infringements Matter**

Let me first take a page to explain my business model and how rampant online theft is destroying it. I've been publishing online since 1995, and the reputation and traffic I built helped me start my own book publishing business while also writing a series of books for McGraw-Hill. My bestselling book for McGraw-Hill, a guide to building PCs, sold over 100,000 copies through four editions and was translated into a half dozen languages. A bestselling how-to book that has to be updated every 18 months barely pays the bills and offers no future security. So I decided to focus on self publishing with on-demand printing.

My publishing business depends on a website to attract new readers, and I typically published around 50% of the material from each new title online. These excerpts have been very popular over the years, by January of 2011, my popular pages for computer troubleshooting and repair were drawing approximately 10,000 visitors a day. The published computer books are among the most top hardware titles on Amazon, used in several college training programs and even by the Department of Homeland Security.

But in recent years, the online work was also drawing undesired attention from black hat SEO (Search Engine Optimization) consultants and plagiarists looking for material to draw visitors to their own websites. Up until the beginning of 2011, Google did a good job keeping track of who was the original publisher of content and awarding that website with higher search rankings than the infringing sites.

In February of 2011, Google attempted to depress the rankings of "content farms", websites that posted stolen content or poorly articles in order to attract visitors to online advertising. These sites have long hidden behind the Digital Millennium Copyright Act (DMCA), claiming that the infringements are posted by community members, and that they are protected by the DMCA Safe Harbor. These same sites syndicate the plagiarized material to hundreds or thousands of other sites, all of whom claim DMCA protection, making it impossible for an author to have all of the infringements removed. These websites profit from the advertising sold, while the people posting the articles gain links back to their websites, to build their own standing in search engines.

Unfortunately, in their February 2011 update, since named "Panda", Google lost track of who was the original publisher of much of the text on the Internet. All of a sudden, Google searches for unique phrase from my books or web pages started ranking web pages that were using text stolen from my work before my own pages in Google search results. By the summer of 2011, Google search was reporting that content from my website was showing up in hundreds of thousands of places.

In some instances, my original pages didn't appear in the results at all, even when telling Google to repeat the search with omitted results included. In other words, the hundreds of thousands of infringements had convinced Google that work on my website was not original, or at least, suspicious enough to penalize. So my Foner Books website that I had spent ten years building lost nearly three quarters of its usual visitors, and my book sales suffered proportionately.

## **Death By A Thousand Cuts**

Some infringements are straight forward copying of entire books or whole web pages. Unfortunately, the deeper problem for my business model is the partial takings and the thin rewrites. While the Google Panda update that was supposed to punish websites with copied material had a devastating impact on my 100% original website, I finally discovered through extensive analysis that some pages that been impacted even more than others. It then became apparent that the worst impacted pages, those that essentially dropped out of the Google results altogether, were also the pages had been attacked professionally, picked apart and republished in sections or with minor inclusions or rewordings.

Picture a web page that is a chapter from a published book which consists of thirty or forty paragraphs. The professional thieves chop the text up into pieces of roughly 200 words, maybe two paragraphs per chunk, and then publish each chunk on a free blog site or a free article site as a stand-alone article, with a unique title tailored to attract search visitors. Then they salt those chunks of text with a couple keywords related to a product they are pushing, often a particular laptop battery. The resulting pages are of no use to human beings, but look spot-on to a search engine that is responding to a search request for that laptop battery. In addition to carrying pay per click advertising, these pages include a link to another website owned by the thief, to build up its search engine reputation.

In the most egregious cases, a single individual created hundreds of unique infringements (ie, posted to different blogs and articles sites) from a page on my website, which was then multiplied to thousands or tens of thousands of infringements by syndication. These professional infringements were limited to a handful of my pages related to laptop troubleshooting and repair due to the monetization possibilities for the people doing the stealing. So while fifty or more of my web pages suffered from large numbers of infringements, Google would report 100,000 or more matches for text from the handful of pages that were most attractive to thieves.

While I can easily identify the websites that benefit from the links embedded in all these infringements, I have no way to prove that the pages were posted by the same people, as strong as the circumstantial evidence is. And since they weren't U.S. based businesses, the idea of making some kind of test case of this in the Federal courts was completely off the table. I don't have the ability to do forensics work on infringements that started in 2009 or 2010, and I don't have the ability to take the fight to China or to Russia.

## **Conclusions**

Independent authors throughout the country are experiencing problems similar to mine, though the severity is closely related to the commercial potential of the work. When I started my Foner Books website back in 2000, the first pages I posted were English translations of my great-grandmother's groundbreaking Hebrew fiction works. Not surprisingly, nobody has ever infringed on those pages. But my bread-and-butter work on computer troubleshooting has attracted huge numbers of infringements, due to its suitability for monetization through Internet advertising, and sadly, it has also been used as a lure to attract people to sites that infect their computers with viruses and malware.

There's nothing anybody can do to prevent copyright infringement on the individual level. The target for copyright law and enforcement mechanisms should be the business of copyright infringement. I'm sure

you'll be receiving plenty of comments about the file sharing networks that enable the distribution of illegal copies of movies, music and books, so I want to concentrate on the less romantic subject of professional SEO (Search Engine Optimization) infringements.

There are two ways for an author to attempt to defend online works. The first is to send DMCA complaints to each and every infringer, and the second is to try to prevent discovery through having the results removed from search engines. There are far too many infringements on my work for me to pursue every website involved. Currently, my focus is almost entirely on the latter, and I've submitted thousands of infringements to Google for removal from their search engine.

If there were a mechanism through which authors and publishers could submit their websites to Google and Microsoft, which runs Bing and Yahoo search, for registration as the authoritative source, I would even be willing to pay a fee for the service. In return, I would expect the search engines to refuse to index web pages from other sites that used substantial amounts of the same text, and to flag as suspicious sites that include a large amounts of duplicated content. What Google attempted last year was to make these decisions without the benefit of inviting publishers to register their content first, and they ended up doing more harm than good.

I understand that the goal of Google and Microsoft in indexing the Internet is to fully automate the process and not get involved in dealing directly with publishers. While this has worked out phenomenally well for their profit margins, it's done unmeasured damage to copyright holders and the people who create the content that make the Internet worth indexing.

The search engines have become the infrastructure of the Internet, the equivalent of the phone company or cable company in the bricks-and-mortar world. Yet Google offers no direct communication path for publishers, no phone number, no e-mail with an individual answer, not even a pay-for-time option. It suits their purposes to pretend that these problems don't exist or are minor until forced to take actions by public outcry or legislation. Intellectual property rights have become inextricably entwined with search engine functionality, and it's time the search engines acknowledge this and start working with the publishing industry to solve these challenges.

Morris Rosenthal