



The Search Fixation: Infringement, Search Results, and Online Content

By Matt Schruers

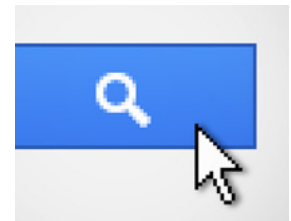
Certain content industry constituencies fixate on search engine results as a mechanism for reducing online copyright infringement. Search engines, it is contended, are the primary tool of copyright infringers seeking to engage in unauthorized downloading or streaming from sites such as the Pirate Bay. The Recording Industry Association of America (RIAA) released a February 2013 report criticizing Google efforts in this area, although this was not the first time that industry stakeholders, pundits, or policy-makers have urged search engines to disappear certain results.

The contention that disappearing undesirable entries from search results would substantially prevent piracy is flawed, however. The solutions to online infringement have little to do with search.

Infringing sites receive limited traffic from search. In the context of music, the available evidence suggests that the frequency with which users input queries like “download,” “mp3,” or “torrent” is relatively low. Generally, search users input fairly broad queries about artists, and these queries lead to a wide variety of lawful sites delivering licensed content. To the extent rights-holders nevertheless desire different results to be displayed in the infrequent cases where users search with terms such as “download,” “mp3,” or “torrent,” however, they and their licensed online distribution partners can do more to improve search engine optimization around these terms.

Background

When drafting the Digital Millennium Copyright Act in 1998, Congress explicitly contemplated search engines as a category of online service which should receive liability limitations to motivate expeditious responses to complaints of third-party online infringement. Section 512(d) of the Act addresses “information location tools,” which include directories, indexes, and search providers. Provided that a service complies with the Act’s “notice and take down” requirements, the Act limits what remedies are available against that service, should it be sued for secondary liability based on infringements occurring on other Internet sites that it has indexed. Upon receiving a DMCA-compliant written notice alleging infringement, search engines – like most online platforms today – expeditiously disable links or access to the infringing content in question. In addition to instituting a DMCA Compliance Program and formulating an “acceptable use policy,” some online platforms develop additional tools or utilities to reduce access to infringing content, where technically feasible. Due to the absence of a comprehensive government database of (a) what works are protected, as exists with patents, and (b) what works have been licensed to whom, online services must rely upon rights-holders or their authorized representatives to provide notice regarding sites making unauthorized uses of content.



Empirical Evidence

Some rights-holder constituencies do not view DMCA compliance and supplemental voluntary efforts as sufficient, and seek to increase the burden of content enforcement for online services. The debate over search results is not new; one of the operative provisions of the controversial “Stop Online Piracy Act” (SOPA) would have obligated search providers and other services to remove or disable access to entire sites upon the allegation of “facilitating” infringing activities.¹ This might lead to the conclusion that search engines are a prominent tool in the infringers’ toolbox.

In fact, available evidence suggests that search is not a particularly relevant tool for infringers seeking to find sites (such as the Pirate Bay) or for sites to find users.

Search results may receive disproportionate attention, however, because they are easily tested.²

Evidence suggests that sites associated with infringement receive relatively little traffic from search. Recent research by BAE Systems Detica found that for most sites associated with infringement, search referral is not a primary source of traffic. Rather, users primarily arrive to these sites directly, e.g., “returning users often type the address directly into the browser,” or via some social networking action.³

Traffic statistics in 2011 indicated that a mere 15% of traffic to alleged “rogue sites” was referred by search,⁴

and current Alexa ratings indicate that only about 8% of the Pirate Bay’s traffic arrives via Google. Moreover, as of August 2013, over 20% of queries that result in traffic being directed to the Pirate Bay consist of words comprising the Pirate Bay’s domain name.⁵ This suggests that users are quite aware of their intended destination before they arrive at a search engine, and that any facilitation was minimal.⁶ Indeed, when Google announced plans to create a new search algorithm signal that would demote sites receiving substantial numbers of DMCA takedowns, prominent piracy sites dismissed the proposal, pointing out that they received little traffic from the search provider.

The Pirate Bay, for example, claimed that “a very low amount of our traffic actually comes from search engines such as google. If you think about it, you’ll understand why – we’re competitors. TPB is a search engine just as Google. We only specialize in different methods and links.”

Isohunt similarly dismissed the move, pointing out that less than a quarter of its traffic came from search, “and much of that being searches on ‘isohunt’ and ‘isohunt.com’”. As a result, Isohunt predicted that it would survive even a complete search engine ban.⁷

Organic Results

In February 2013, an RIAA paper complained that, notwithstanding these steps, particular search results were not being sufficiently demoted.⁸ The basis for the complaint was a number of searches for “[artist name] [track name]”, appended with either “download”

¹Brian T. Yeh, *Online Copyright Infringement and Counterfeiting: Legislation in the 112th Congress*, Congressional Research Service (Dec. 2011).

²This type of observer bias has been described by some social scientists as the “lamp post effect.” David Freedman, *Why Scientific Studies Are So Often Wrong: The Streetlight Effect*, Discover Magazine, Dec. 10, 2010, available at <http://discovermagazine.com/2010/jul-aug/29-why-scientific-studies-often-wrong-streetlight-effect>.

³BAE Systems Detica, *The Six Business Models for Copyright Infringement*, June 2012, at 5, available at <http://www.prsformusic.com/aboutus/policyandresearch/researchandconomics/Documents/TheSixBusinessModelsofCopyrightInfringement.pdf>

⁴Mike Masnick, *Data Shows: Removing ‘Rogue Sites’ From Search Won’t Make Much Of A Difference*, Techdirt, Nov. 30, 2011, available at <http://www.techdirt.com/articles/20111130/05022316931/data-shows-removing-rogue-sites-search-wont-make-much-difference.shtml>

⁵See Alexa Web Information, available at <http://www.alexa.com/siteinfo/thepiratebay.sx>. The site has moved from the .org TLD to several different ccTLDs since 2012.

⁶Masnick, *supra* n.4.

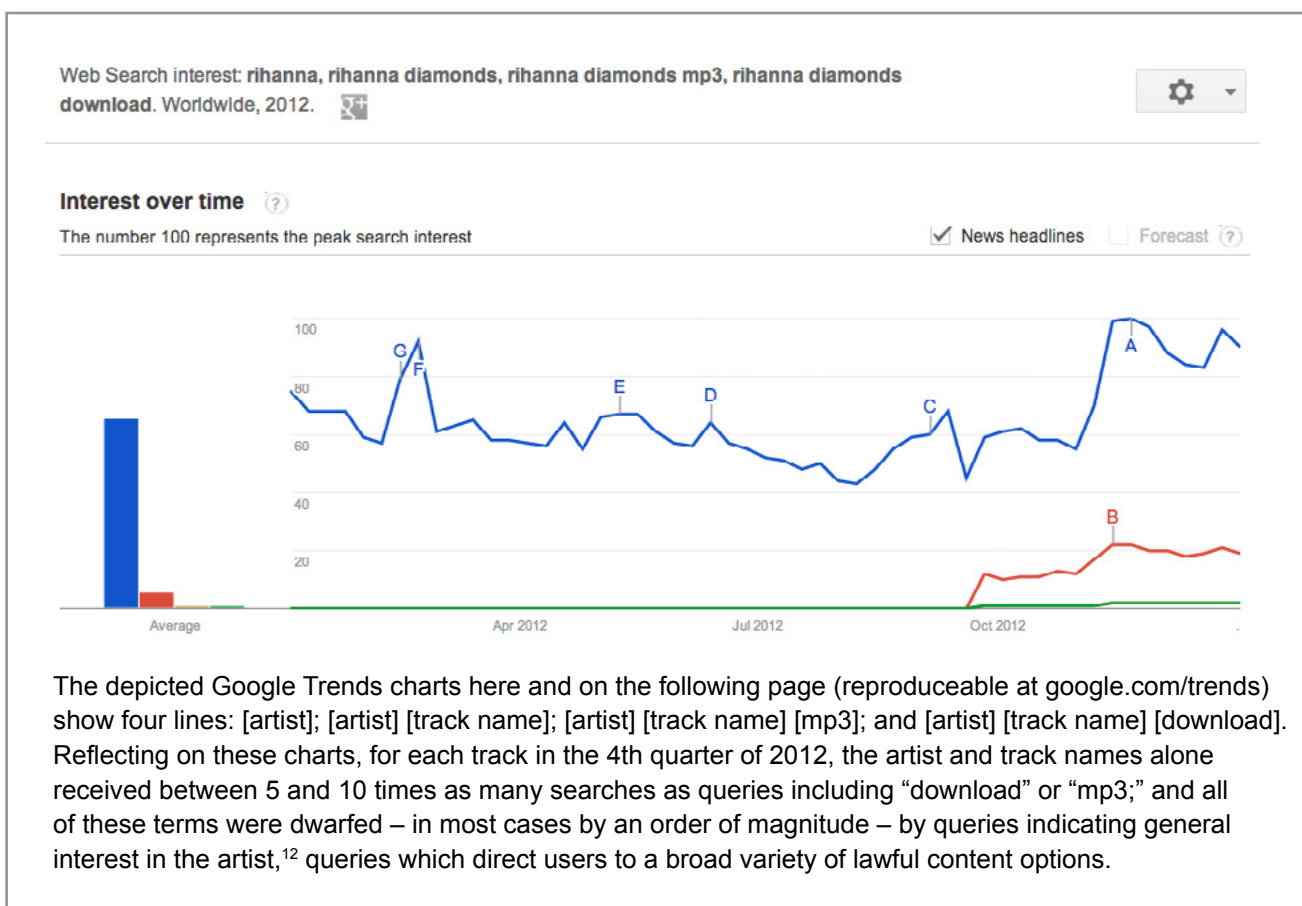
⁷The Pirate Bay, *Google ranking TPB (and others) lower*, Aug. 15, 2012, available at <http://thepiratebay.sx/blog/220>; see also Digital Digest, *The Pirate Bay, isoHunt Responds To Google DMCA Demotions*, Aug. 20, 2012, available at <http://www.digital-digest.com/news-63452-The-Pirate-Bay-isoHunt-Responds-To-Google-DMCA-Demotions.html>.

⁸Available online at https://www.riaa.com/blog.php?content_selector=riaa-news-blog&blog_selector=Googles-Move-&blog_type=&news_month_filter=2&news_year_filter=2013

or “mp3.” Although some users did appear to observe the effects of demotion,⁹ it is possible that certain, popular sites will retain high search ratings. Most modern search engines rely on numerous cues, and the quantity of DMCA notices will only be one among many.¹⁰

Concerns about organic search results containing terms such as “mp3” or “download” are misplaced, however. Actual search data indicates that appending “mp3” or “download” as the RIAA paper suggests is

statistically uncommon. Users far more frequently search for “[artist]” or “[artist] [track]”. Consider, for example, the paper’s leading examples: Rihanna’s “Diamonds” and Kesha’s “Die Young.” Google Trends data indicate that only a small fraction of searches for the artist’s name and track name also included the words “mp3” or “download.” Similarly, notwithstanding the international sensation surrounding South Korean musician Psy’s “Gangnam Style,” only a small fraction of search results included the suspect terms “download” or “mp3.”¹¹




⁹See TorrentFreak, *Google Downranks The Pirate Bay in UK Search Results*, March 6, 2013, available at <http://torrentfreak.com/google-downranks-the-pirate-bay-in-uk-search-results-130306/>

¹⁰For the site mp3skull.com, which figures prominently in the RIAA’s complaint, less than 5% of the content has been the subject of a complaint. See Google Transparency Report, available at <http://www.google.com/transparencyreport/removals/copyright/domains/mp3skull.com/>. Over-prioritizing this variable could have unfortunate effects when abusive DMCA notices are issued – such as the recent DMCA complaint by CNBC after one of its anchors fared poorly in a debate with Sen. Elizabeth Warren, and the Senator placed the video on her site (clearly, supported by a strong fair use argument).

¹¹It could be a topic for future research to explore whether the use of “mp3” or “download” in queries correlates in any way commercial availability of a given work across regions.

¹²Google’s more recent practice of providing “Instant” results (not to be confused with autocompleting queries) may further diminish the relevance of terms like “mp3” and “download,” since a user searching for “Psy Gangnam Style” will first see likely responsive results for “Psy” and then “Psy Gangnam,” before appending any other terms – thus decreasing the likelihood of the user navigating to undesirable results.

Web Search interest: **kesha, kesha die young, kesha die young download, kesha die young mp3**. Worldwide, 2012. 

Interest over time 

The number 100 represents the peak search interest

News headlines Forecast 

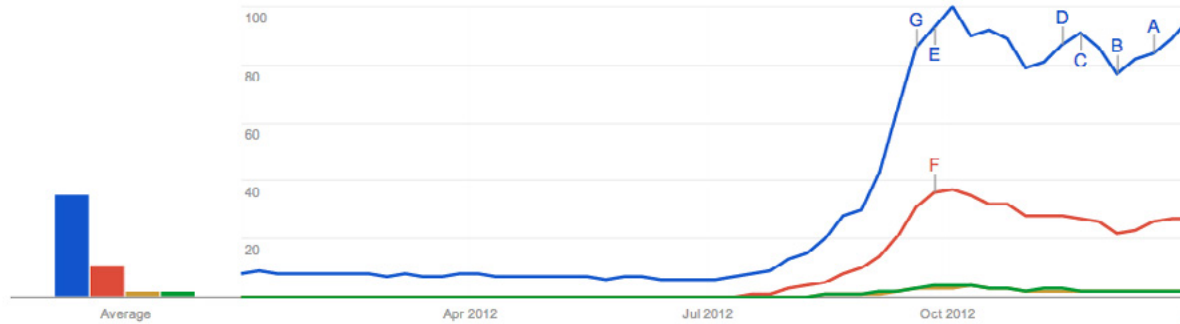


Web Search interest: **psy, psy gangnam style, psy gangnam style download, psy gangnam style mp3**. Worldwide, 2012. 

Interest over time 

The number 100 represents the peak search interest

News headlines Forecast 



Increasing the Visibility of Lawful Content in Search Results

In addition to being of little practical value, demoting sites based on perceived lawfulness poses numerous challenges. If “relevance” cannot be used, what metric should determine which results are displayed first for “Psy Gangnam Style” or “Madonna MDNA”? Different rights-holders will likely have different opinions on this matter.

Recently released research suggests that more effective solutions exist. A July 2013 study by music streaming service Spotify illustrates that viable commercial options for online content have been effective at diminishing online piracy.¹³ Additionally, research recently published in Norway concluded that the arrival of Netflix and Spotify in that nation was followed by an 80% reduction in music piracy and a 50% reduction in video piracy.¹⁴

The fixation on demoting responsive but undesirable search results overlooks a more viable strategy: promoting desirable search results. In general, Internet discovery is a tool that greatly aids artists. For example, research has found that the total quantity of new albums released annually has increased sharply since 2000, driven by independent labels and purely digital products. This coincides with decreased concentration of sales in the top albums, such that increasing numbers of albums find commercial success even when they do not receive substantial broadcast radio airplay.¹⁵ The number of authorized music services selling this content increased five-fold between 2007 and 2011.¹⁶ As a result, artists are no longer consigned to finding favor with a limited group of gatekeepers. The explosion of success around the musician Baauer’s “Harlem Shake,” which sold nearly one million digital downloads, for example, was driven entirely by YouTube videos.

Many music sites now demonstrate an acute awareness of the importance of a strong digital presence, and generally demonstrate effective organic and paid search optimization. Searches for such terms as “music downloads” indicate that lawful platforms such as Spotify, Last.fm, and Rdio aggressively seek to optimize their organic (i.e., “natural”) search results, as well as paid search advertising for such terms, including terms that might otherwise lead to unlicensed sites.

1. Licensee Optimization

Nevertheless, lawful services appear not to be optimizing their sites to appear as results for the queries in the RIAA paper – “mp3” and “download.” This suggests either (a) a deficiency in otherwise robust online marketing strategies, or (b) that these terms are judged to be unworthy of optimizing because they will drive a trivial amount of commercial traffic. Stated otherwise: if search terms such as “mp3” and “download” were likely to lead to sales or subscriptions, a rational, profit-minded online platform engaging in basic search engine optimization (SEO) would attempt to incorporate those terms in site content. The failure to do so appears to reinforce the conclusion from the Trends data: these searches are not economically consequential.

Regardless, licensors can negotiate with online services to include such terms in site content. At present, a lawful commercial site is unlikely to appear in organic search results for a query including “download” or “mp3” if those terms do not actually appear in the indexed pages of the site.

For example, while Netflix aggressively promotes its content in paid search advertising, at publication of this paper the site’s robot.txt file prevents search indexing of the specific page for the hit program,

¹³Will Page, “Adventures in the Netherlands: Spotify, Piracy and the new Dutch experience,” July 2013, available at <http://press.spotify.com/us/2013/07/17/adventures-in-netherlands/>

¹⁴Sophie Curtis, *Spotify and Netflix Curb Music and Film Piracy*, The Telegraph, July 18, 2013, available at <http://www.telegraph.co.uk/technology/news/10187400/Spotify-and-Netflix-curb-music-and-film-piracy.html>.

¹⁵Joel Waldfogel, “And the Bands Played On: Digital Disintermediation and the Quality of New Recorded Music” (Univ. Minnesota, NBER 2012 (prelim. draft)) available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2117372.

¹⁶Michael Masnick & Michael Ho, “The Sky is Rising” (Floor64 2012), at www.techdirt.com/skyisrising/.

“Breaking Bad.”¹⁷ The fact that Netflix’s “Breaking Bad” description page content goes unindexed may impair the search ranking for that show, from that platform. As a result, this lawful, affordable, and consumer-friendly option may not appear as high in search results as it could. To address situations like this, licensors could contract with online platforms to ensure greater indexing, and utilize keywords in site content through a standard term in a licensing agreement.

In addition, a licensor could request, for example, that sites include language to the effect of “accessing [artist] [track] from unauthorized mp3 or torrent download sites, or streaming sites may violate federal law.” By doing so, lawful sites may then be included as potentially responsive organic results to queries including the terms “mp3,” “download,” and “torrent.” This is not a novel proposal; the National Association of Recording Merchandisers suggests such strategic optimization in an infographic released in May 2013, urging music sites to “direct users away from torrent sites” by using torrent as a keyword in site content preparation.¹⁸

Some value is provided by websites such as “Why Music Matters”¹⁹ and the MPAA’s recently launched website, “Where to Watch.Org” – both are portals offering indexes of authorized content platforms. These portals will improve the search ranking of the sites to which they point, but because the portal content does not advertise specific media (in other words, wheretowatch.org is highly unlikely to appear when a user searches in Bing, Yahoo, or Google specifically for “breaking bad amc episodes”), these portals are unlikely to appear as results for queries involving specific media. Thus, NARM also recommends that licensors “link to stores and apps that sell your music” and “leverage social media for creating backlinks to increase search engine visibility.”

2. Licensor Optimization

Just as rights-holders can use the licensing process to ensure licensees engage in more aggressive SEO, they can also take action on their own side. Some licensors now promote their own licensees’ content by linking to licensees in relation to the content licensed. Thus, for example, the official Rihanna website points to iTunes, but not other platforms from which that content is available (as of publication). However, as of publication, the term “torrent” appears nowhere on the rihannanow.com domain. As a result, the artist’s official site is unlikely to be served as a responsive result to a hypothetical search like “rihanna diamonds torrent.” The inclusion of NARM-recommended text (“don’t torrent; buy [here] instead”) on the artist’s site would remedy this. Linking users to other lawful music services in addition to iTunes could be another way to contribute toward improving those services’ page rank. Similarly, Universal Music points to Rihanna’s official website, Twitter account, and Facebook page, but points to no commercial websites from which the artist’s music is available. Addressing this would also contribute toward improved page rank.

Conclusion

While DMCA notices and DMCA compliance programs are one component promoting a robust digital marketplace, efforts to disappear search results are unlikely to mitigate online infringement, in large part due to the irrelevance of general-purpose search engines in the average infringer’s toolbox. A more robust strategy would entail licensors and their licensees focusing on strategic search engine optimization – including but not limited to ‘objectionable’ terms – so as to promote the page rank of lawful sites and increase the visibility of legitimate online content offerings.

¹⁷David Rothschild, “Why is Netflix Hiding Their Non-Original Content from Google?”, May 29, 2013, *available at* <http://insightforge.com/2013/05/netflix-robots-txt/>.

¹⁸Infographic, NARM, “Search Engine Optimization for Music Websites,” May 6, 2013, *available at* <http://www.narm.com/news/infographic-search-engine-optimization-for-music-websites/>

¹⁹<http://www.whymusicmatters.org/find-music>