

# Understanding the Role of Search in Online Piracy

Prepared by Millward Brown Digital for the MPAA



---

# Introduction

The Motion Picture Association of America (MPAA) commissioned Compete to assess the role search plays in how US and UK consumers find copyright infringing copies of TV and film content on the internet. To accomplish the goal of the study, Compete used its online consumer panel to investigate how search is used in the discovery and navigational journey to these infringing content files located across the web. Compete also analyzed whether the algorithm change implemented by Google in August 2012, which incorporated copyright notice levels into its search engine ranking, caused any change in the role that search plays in access to copyright infringing content. The analysis included in this report is based on US and UK consumer data.

---

# Executive Summary

The study identifies that:

- Overall, search engines influenced 20% of the sessions in which consumers accessed infringing TV or film content online between 2010 and 2012.<sup>1</sup>
- Search is an important resource for consumers when they seek new content online, especially for the first time. 74% of consumers surveyed cited using a search engine as either a discovery or navigational tool in their initial viewing sessions on domains with infringing content.
- Consumers who view infringing TV or film content for the first time online are more than twice as likely to use a search engine in their navigation path as repeat visitors.
- The majority of search queries that lead to consumers viewing infringing film or TV content do not contain keywords that indicate specific intent to view this content illegally. 58% of queries that consumers use prior to viewing infringing content contain generic or title-specific keywords only, indicating that consumers who may not explicitly intend to watch the content illegally ultimately do so online. Additionally, searchers are more likely to rely on generic or title-specific keywords in their first visit than on subsequent visits.
- For the infringing film and TV content URLs measured, the largest share of search queries that lead to these URLs (82%) came from the largest search engine, Google.
- The share of referral traffic from Google to sites included in the Google Transparency Report remained flat in the three months following the implementation of Google's "signal demotion" algorithm in August 2012.

---

<sup>1</sup> Methodological detail: Both the *relevance* and *recency* of search queries were used to calculate this figure, versus solely attributing influence to the domains that a consumer visited immediately prior to viewing infringing content. This research methodology was used because consumer navigation to infringing content is complex and could include multiple search queries and visits to infringing websites before a user is able to find a working URL. Therefore direct referrals do not fully reflect consumers' navigation paths.

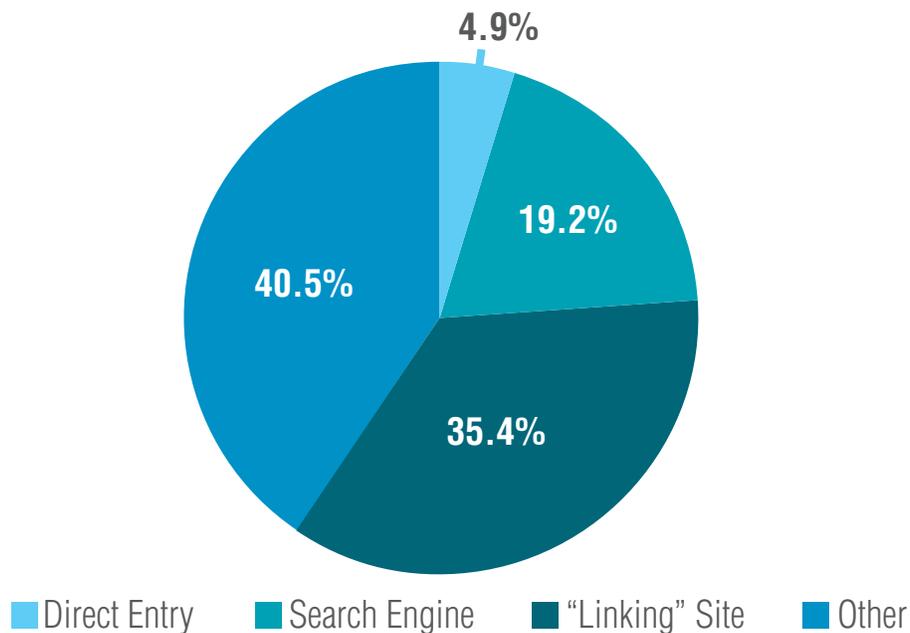
# Analysis

## The Role of Search

The goal for this study was to understand the role that search engines play in the discovery of and navigation to TV and film content online. Specifically, the study aimed to better understand the online paths consumers utilize when they find and view infringing content. As highlighted in the Methodology section at the end of the study, Compete used a hybrid approach to define the role of search, incorporating both the relevancy of keywords used and the recency of the query in relation to when the infringing content was viewed. The approach established instances of consumers visiting infringing content as endpoints, and then attributed search queries that included a related term and also occurred within a twenty-minute prior window as having influenced the path during which the infringing viewing behavior occurred. This methodology highlights the prevalence of consumers using search engines for content discovery. However, this method did not seek to indicate the degree to which infringing content appears on search engine results pages themselves.

Using this approach, Compete observed that on average, approximately 20% of all visits to infringing content were influenced by a search query from 2010-2012. The level was relatively flat between 2011 and 2012 and up from 2010 (16%). Meanwhile, 5% of sessions were driven by consumers who directly navigated to the website where the content is hosted. 35% of viewing sessions were referred by a linking site such as tv-links.eu, and 41% of these infringing viewing sessions resulted when a user clicked a link on any other type of website like a social network, forum, and blog or used a bookmark.

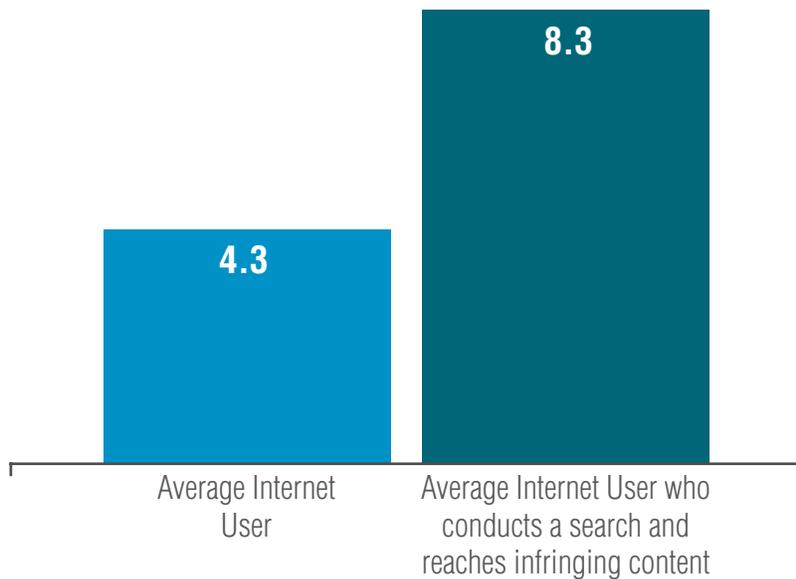
**Share of Visits to Infringing Content by Entry Method**  
(% of Visits to Infringing Content URLs, 2010-2012 Monthly Average)



## Search Behavior: How do searchers behave?

Consumers who use search as part of the path to viewing infringing TV or film content are heavier search users compared to the average internet user. These individuals conduct an average of 8.3 searches during their viewing session versus 4.3 searches for average internet users, equating to 2X more searches during sessions when infringing behavior occurred. While this analysis indicates that search behavior intensifies among consumers who view infringing content, it is also likely that this segment of consumers and this specific online activity are naturally prone to intensified search behavior since online video consumption is a more sophisticated behavior relative to simpler weather or news update searches.

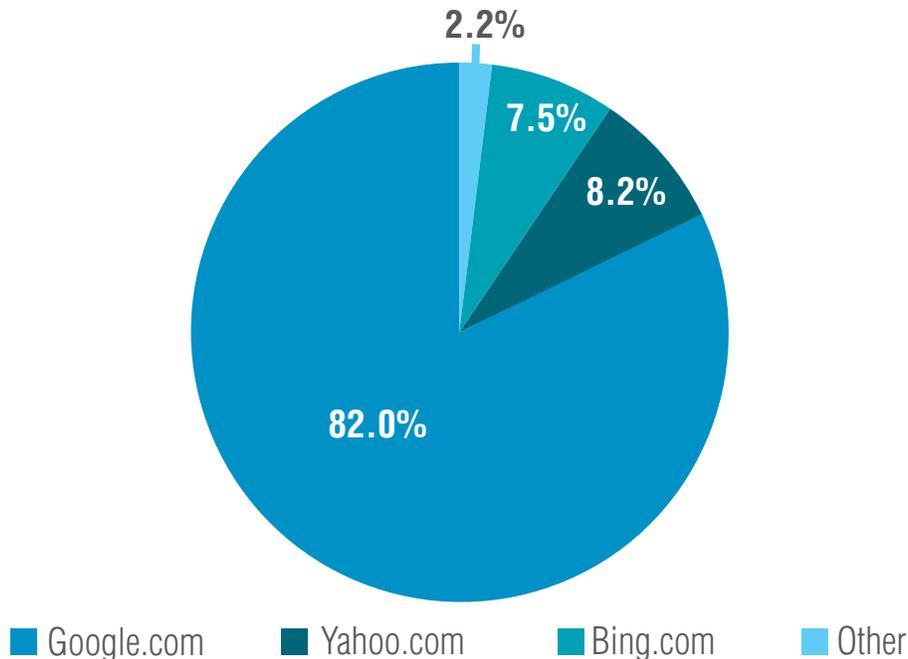
**Volume of Search Queries per Session**  
(Number of Queries per Session, 2010-2012 Monthly Average)



For consumers who conducted a search prior to viewing the infringing content analyzed, 82% navigated through to the content via Google, compared to 16% from Yahoo/Bing and 2% from other engines. Approximately half of consumers that used search to reach infringing content navigated to this content within two to seven minutes, and just under 10% reached the infringing content in less than a minute after conducting a search.

## Search Engine Share

(% of Search Influenced Visits to Infringing Content URLs Driven by Each Major Search Engine, 2010-2012 Monthly Average)



## Search Term Analysis: What queries are consumers using to reach infringing content?

There are multiple ways consumers use search engines and reach infringing film and TV content online. This especially depends on the level of knowledge a given searcher has of the infringing ecosystem. This study analyzed the specific keywords used by consumers to reach infringing content, and classified them in the following categories:

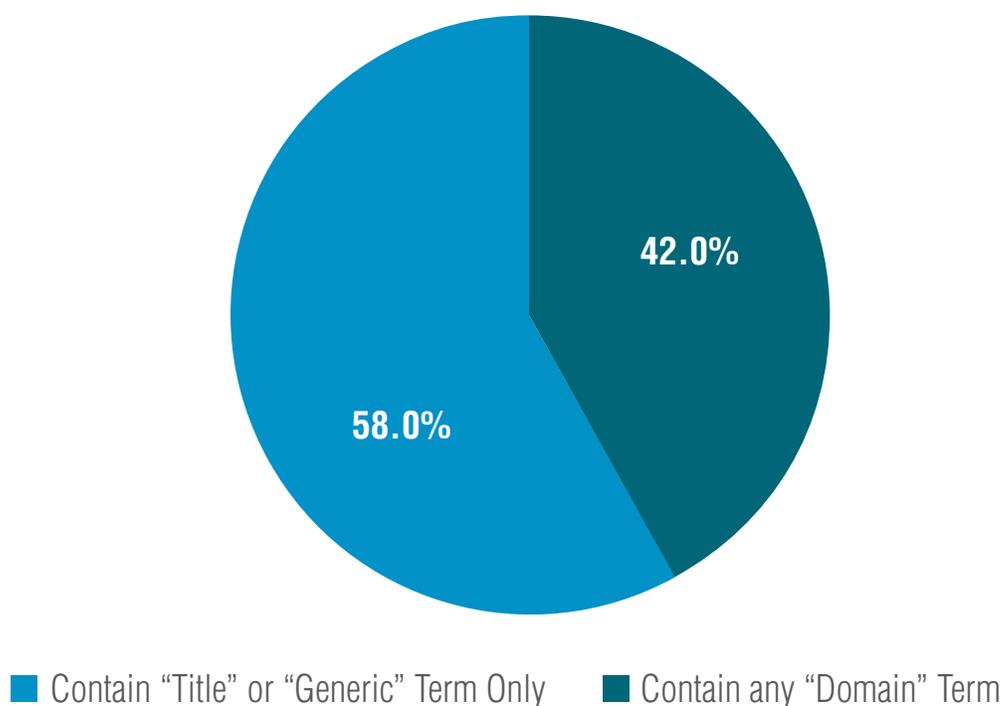
| SEARCH TERM CATEGORY | DEFINITION   | EXAMPLES                                 |
|----------------------|--|--|
| <b>PIRACY DOMAIN</b> | Searches that contain specific domains that are known to host or link to infringing content online | 1Channel, Megavideo, tv-links, etc.      |
| <b>TITLE</b>         | Searches that contain titles of recent TV shows and films  | Lost, Dark Knight Rises, Game of Thrones |
| <b>GENERIC</b>       | Searches that contain phrases related to watching film/TV online                                   | watch tv online, free movies, etc.       |

Compete analyzed the role that “Generic Piracy” terms like “cam”, “divx”, “rip”, etc. play in the consumer journey but found that it was not a significant category and did not warrant being broken out separately.

When looking at keyword types, the majority (58%) of searches by consumers who reach infringing URLs contain only “Generic” or “Title” keywords. These searches do not include specific infringing content websites, indicating that these consumers did not display an intention of viewing content illegally. Meanwhile, 42% of searches contain a “Domain” term, indicating that these users know which domains they intend to go to stream or download the content they are interested in. More granular analysis shows that 37% of searches are “Domain Keyword Only” implying that consumers are leveraging search as a navigational tool by inputting the domain name into the search box or URL bar instead of simply going to the site directly. While “Title Keyword Only” searches (8.8%) are not a primary driver to infringing content by themselves, consumers who used these terms (and may not have originally intended to find this type of content) subsequently navigated to a domain where they could view infringing material.

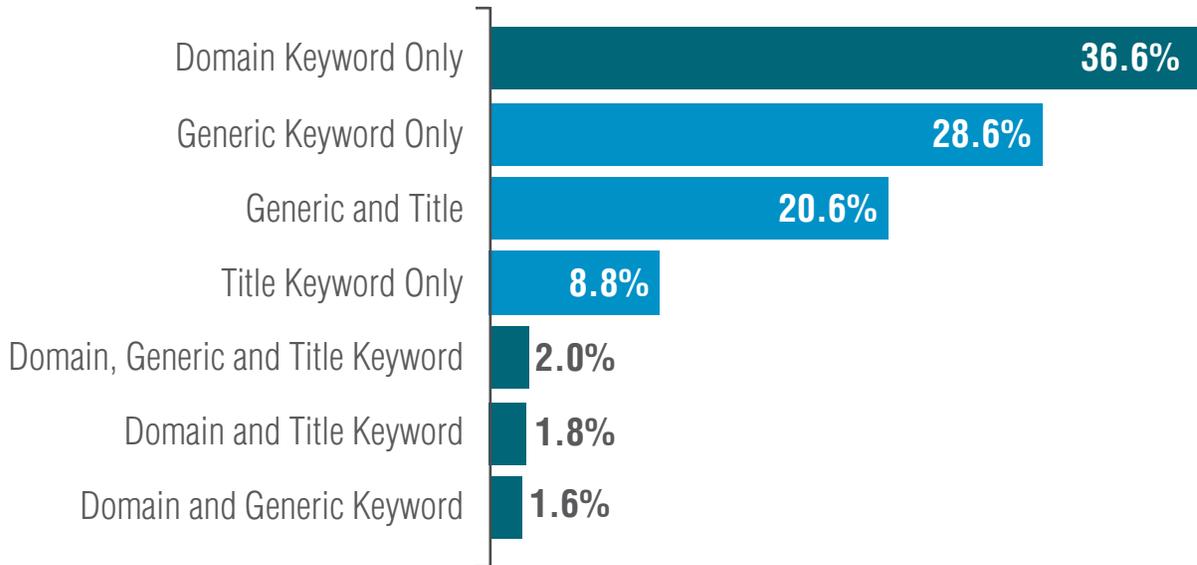
### Search Term Keyword Type Distribution

(% of Visits to Infringing Content Referred by Search by Each Search Term Type, 3 Year Average, 2010-2012)



## Search Term Keyword Breakdown

(% of Visits to Infringing Content Referred by Search by Each Search Term Bucket, 3 Year Average, 2010-2012)

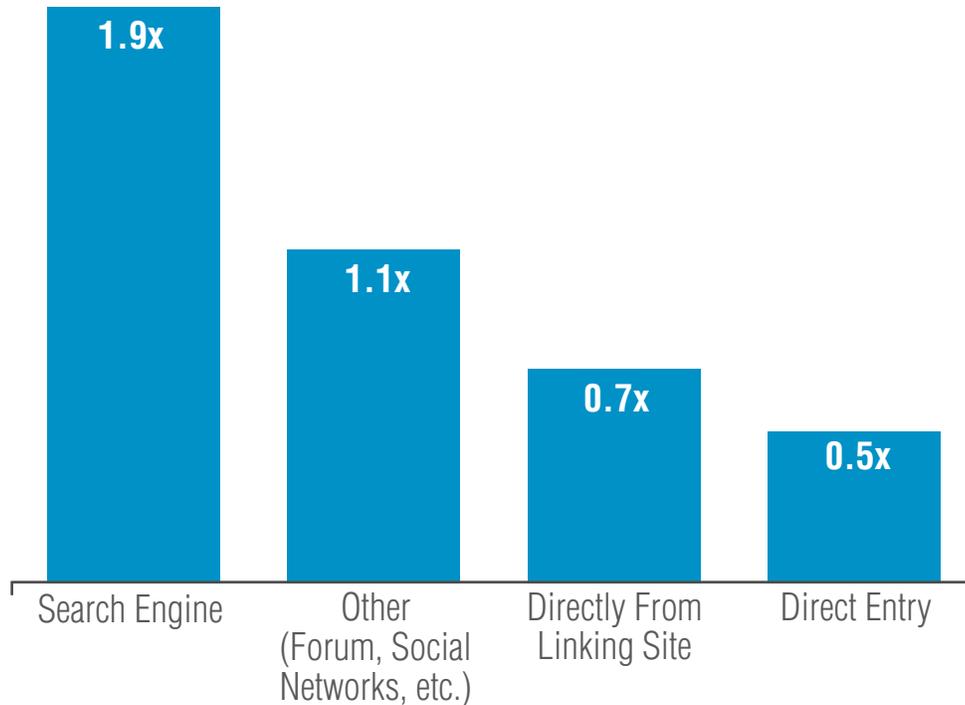


### First Time vs. Repeat Infringing Viewing: Does the Role of Search Change?

One of the biggest goals of Compete's research was to understand the different role that search plays for first-time visitors to infringing domains compared to repeat visitors. Do first-time viewers leverage search more heavily in order to locate a site where they can stream or download content and then navigate directly to that site in the future? It is important to isolate these two segments as looking at an average rate of search influence could potentially be understating the real influence search engines play in this area. Repeat visitors could have been influenced by search at one point, but no longer use it. In our view the initial role of search needs to be assessed. To accomplish this goal, Compete used its longitudinal clickstream panel to segment consumers whose online activity was present in three consecutive months into two groups – First Time Viewers and Repeat Viewers.

## First Time Views vs. Repeat Viewers

(Lift to each entry method among First Time Visits vs. Repeat Visits, 2010-2012 Monthly Average)

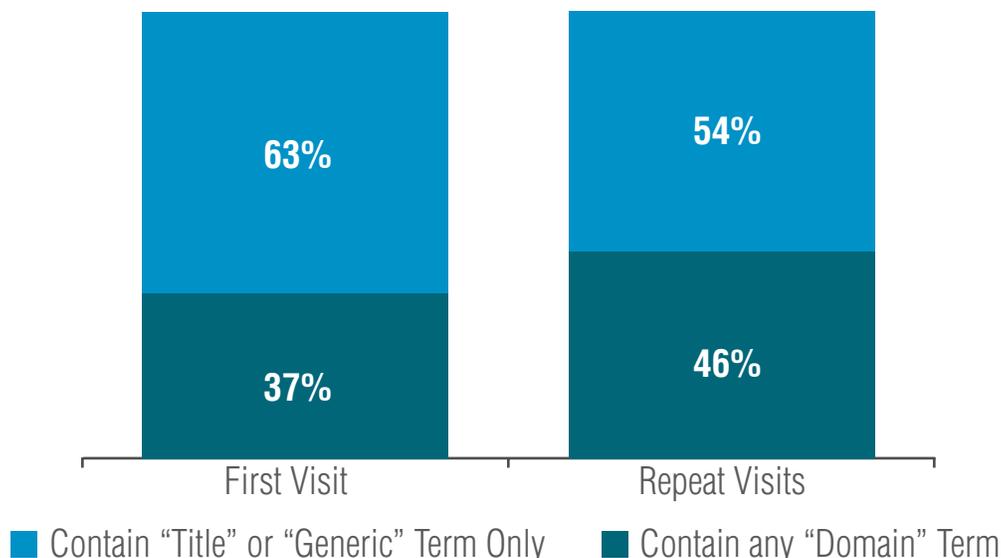


First-time visitors to infringing content were almost 2X more likely to use search on their first visit to infringing content compared to repeat visitors. This confirms that search is principally used for content discovery and is relatively more valuable to consumers who do not know where this content can be found online. This is not surprising and is consistent with other categories of search behaviors. As consumers become aware of content sites, search may play a less direct role than it did during their initial “discovery” session.

In addition to analyzing the clickstream patterns of infringing content viewers, Compete also surveyed consumers to better understand how they discover and navigate to this content and whether search influences their path. Attitudinal responses mirrored the behavioral data; first time visitors indicated that they were 1.6X more likely to use search on their first visit than on repeat visits. 74% of respondents reported that search played a role the first time they visited an infringing content domain, either as a tool for discovery (41%) or navigation (33%), compared to 47% of survey respondents who cited search as being used during later viewing sessions. Note that these responses were not constrained by the five-month window of the observed clickstream data, and represent the respondents’ recollection of their first visit.

## Search Term Keyword Type Distribution

(% of Visits to Infringing Content Referred by Search by Each Search Term Type, First Time vs Repeat Visits, 3 Year Average, 2010-2012)



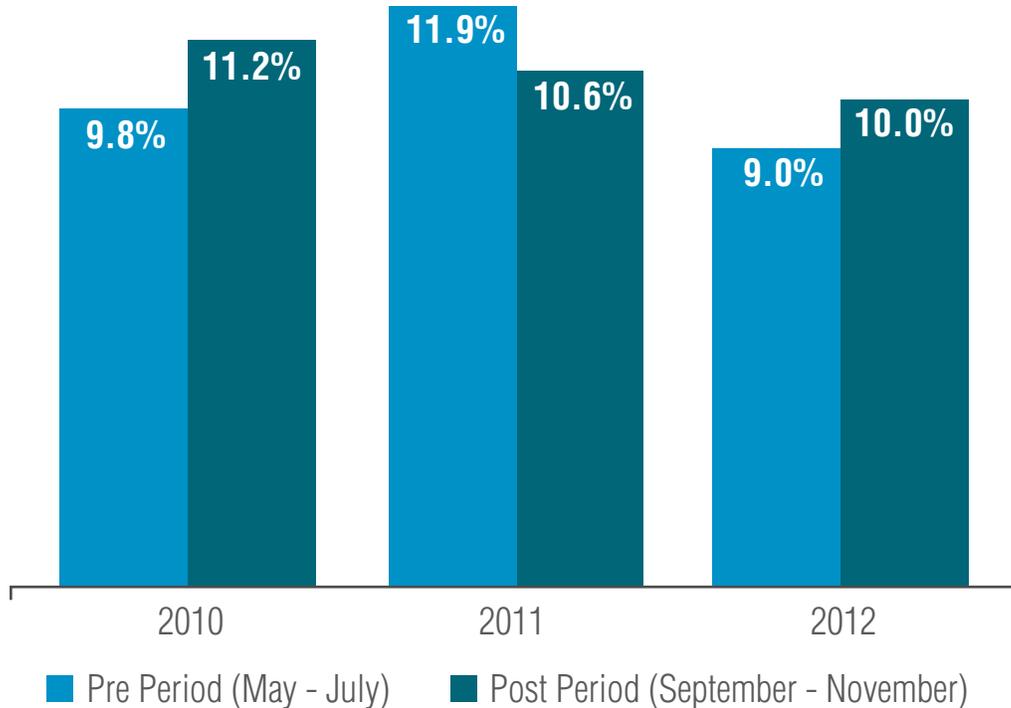
Compete also analyzed if the actual search terms used by consumers differed when comparing first time vs. repeat visitation to infringing content URLs. Analysis showed that first-time visitors were more likely to use generic or title-only keyword terms than repeat visitors (63% vs. 54%). Conversely, repeat visitors were more likely to use domain-specific search terms such as "1channel" or "watch-free-movies.com" on repeat visits at a rate of 46% compared to 37% among first-time visitors. This suggests that repeat visitors to infringing content often utilize "navigational" queries on search engines to arrive at domains where they have viewed infringing content in the past.

## Google Transparency Report: Has it changed anything?

The study also analyzed the extent to which the "signal demotion" algorithm change implemented by Google in late 2012 impacted the role of search in consumer discovery and navigation to infringing content domains online. The stated goal of the algorithm change was to demote or remove search listings that were reported as hosting or linking to infringing content based on submissions sent to Google from content owners.

Other industry research has similarly analyzed the impact this algorithm change had on the infringing content ecosystem, but mostly from a "supply" perspective. These prior studies measured the presence and listing rank of various infringing sites on Google search result pages over time to identify any impact based on the volume of notices sent to Google for these same domains. Compete analyzed its clickstream panel to measure the consumer demand side of the Google change. This analysis sought to identify whether the number of consumers reaching infringing content sites from Google changed after the algorithm was released.

**Google Transparency Report Site Visitation - Google Search Referred**  
(% of Visits to GTR Sites Referred by a Google Search, Pre vs. Post, 2010-2012)



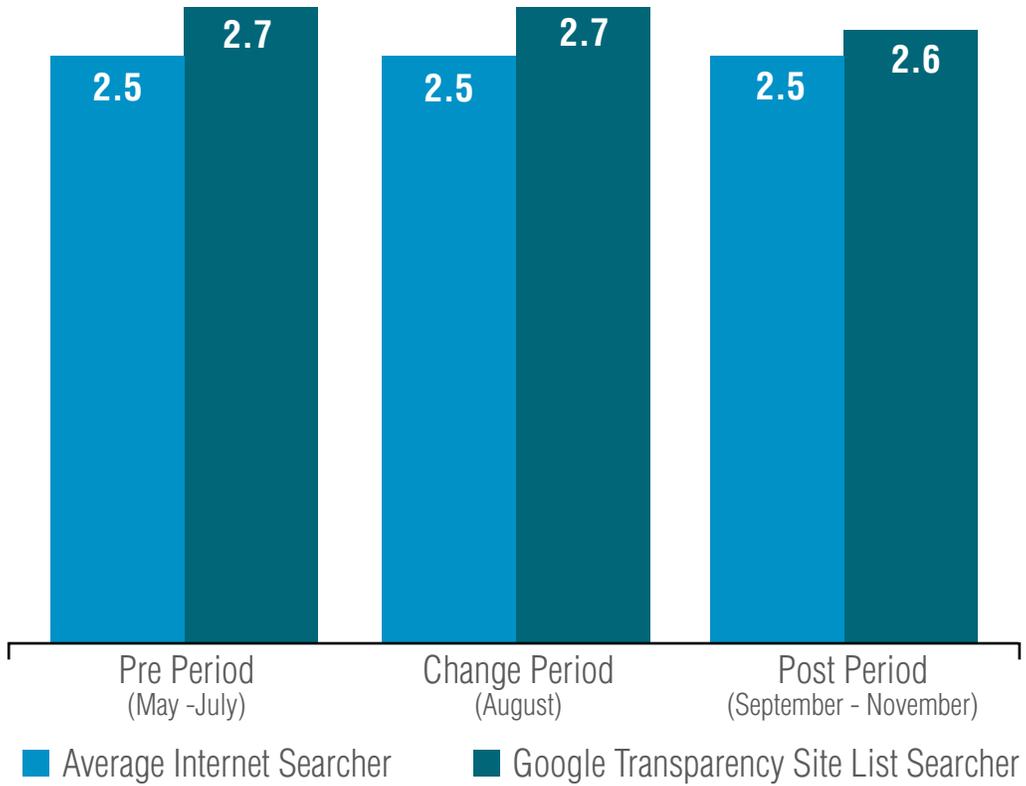
For the two years prior to the study, Google directly referred between 8% and 12% of traffic to the sites analyzed in the Google Transparency Report (“GTR sites”). When comparing the three months before and after the algorithm was implemented in 2012, the share of direct referrals from Google to these sites increased slightly from 9% to almost 10%. Also, in comparison to the same periods from prior years (to isolate seasonality), the data indicate there was not a statistically significant change in direct referrals from Google to the GTR sites during the three months following the implementation of the algorithm.

In addition to analyzing referral volume, the study also measured the impact of the algorithm on consumer behavior. Specifically, the analysis gauged the extent to which consumers were required to find formerly higher-ranking results lower on search engines results pages or on subsequent pages due to the algorithm. This research indicates that among consumers who navigated from a Google search result page directly to a site listed on the GTR, the average listing placement decreased slightly from 2.7 to 2.6 when comparing the three months prior to and following the implementation of the algorithm (in this instance, a lower number translates into a higher ranking on search engine results pages, not a lower ranking). Therefore, the data do not indicate a significant change in listing placement; the search results consumers accessed were not lower in ranking than prior to the algorithm change.



### Google SERP Clickthrough Placement

(Average Placement on Page Where Searcher Actually Clicks, Organic Listings Only)



---

# Methodology

This study was based on the analysis of a database of 12 million film and TV content URLs that were known to host infringing content over the period 2010-2012. The URLs referred to the piece of content, not the site domain. The database was provided by two internet scanning vendors (DTecNet and IP Echelon) which on a regular basis scan websites for a list of MPAA member company titles, in order to find active infringing links and determine to which host sites the links direct. Note that these URLs are website-based and therefore the analysis based on these URLs does not include P2P sites or applications.

Using its robust clickstream panels in the U.S. and U.K. (2M and 200,000 members, respectively), Compete analyzed how many US and UK internet users reached the infringing URLs, how often and how long they visited for and most importantly, how they got there. In this study, Compete focused on the **role that search plays** in the consumer path to infringing content. Compete also identified other ways that consumers arrive to infringing content including by either directly going to an infringing domain or URL or by reaching that content via a “Linking” site like tv-links.eu. Compete’s data was used to identify aggregate trends across different consumer segments and was not used to report individual infringing panelists.

## Defining the Role of Search

Compete is able to measure how consumers use search engines in their online navigation process and is able to isolate the specific engine, term, what was displayed on a user’s screen and what link they clicked. Consumers use search in two ways within the online piracy ecosystem: consumers can use search as a way to **discover websites** that link to or directly host the content they are looking for, or it can be used as a **navigational tool** to arrive at a known linking or hosting website instead of typing the URL into the browser.

Compete developed a customized, hybrid approach for this study to measure the impact that search plays in the path to infringing content as accurately as possible. This holistic approach contrasts with a more narrow definition that counts search only when a visit is preceded by a visit to a search engine. Compete defined a “search referral” to infringing content by:

- **Keyword Centric Component: A user must have conducted a relevant search query.** “Qualifying” search queries include all types of phrases that Compete has observed which online consumers use to find illegal TV and film content online. They include domain terms like “1Channel” and “sidereel”, generic terms like “watch movies online” and movie and TV title-based terms like “Dark Knight Rises”..
- **Time Centric Component:** In the same session (defined as a continuous period of internet activity without a 30 minute lapse) that a user visited an infringing content URL, users must have conducted a “qualifying” search on any major search engine within a 20 minute window before reaching that same infringing content URL.
  - » In order to capture the most relevant search queries yet also minimize the amount of noise in the data, Compete capped the time window at 20 minutes. Compete’s analysis showed 20 minutes represents the most appropriate cutoff time as terms were still relevant to the content reached. After the 20 minute mark, Compete identified a sharp drop in level of relevancy the keywords had with watching or downloading content onlinewatching or downloading content online.

---

## Defining First Time vs. Repeat Infringing Viewers

One of the key components of the analysis completed by Compete was to understand if search plays a different role among consumers who are reaching infringing content on a site for the first time compared to those who have visited the site before. In order to accomplish this task, Compete segmented panelists who were in its panel for five consecutive months between August 2012 and December 2012 in order to perform a longitudinal analysis with a sufficient sample. The first visit to an infringing content domain within the time period was denoted as a “first-time” visit and all subsequent visits to that domain in that window were denoted as being “repeat” visits. Once the segmentation was defined, Compete was able to analyze the role that search played among the same group of consumers who visited infringing content at various times.

## Survey Methodology

In tandem with its clickstream analysis, Compete conducted a quantitative survey to uncover additional learnings regarding how consumers use search in the steps leading up to watching or downloading infringing content. Additionally, Compete leveraged the survey to better understand the role that search plays for first-time infringing content consumers versus repeat visitors. Via email and an online survey, Compete interviewed 565 US and 500 UK panelists aged 18-54 who visited specific infringing URLs as well as known infringing domains in 2012. The survey was conducted between December 15th 2012 and January 7th 2013. Survey data was weighted on age and gender to represent the observed online infringing viewing population in each country.

## Google Transparency Report Methodology

In order to understand if Google’s “signal demotion” algorithm change had any impact on how consumers were reaching infringing sites, Compete retrieved the top 2,000 websites from the Google Transparency Report (GTR) website ranked by number of removal requests. We then isolated the share of traffic to these sites that were directly referred via a Google search result page before the change went into effect in August 2012 compared to afterwards in order to ascertain if the change had any effect. The pre-period of analysis consisted of the three months leading up to the change; May, June and July. The post-period was the three months following the change; September, October and November.

---

## About Compete

Compete, Inc. is a Millward Brown Digital company based in Boston, MA. The company is a research and consulting firm that maintains an online behavioral database comprised of 2 million US internet users and 200,000 UK internet users. Compete uses its data to help brands, publishers and agencies understand how people use the web to communicate, shop, research, and consume content and media.

## About MPAA

The Motion Picture Association of America, Inc. (MPAA), together with the Motion Picture Association (MPA) and MPAA's other subsidiaries and affiliates, serves as the voice and advocate of the American motion picture, home video and television industries in the United States and around the world. MPAA's members are the six major U.S. motion picture studios: Walt Disney Studios Motion Pictures; Paramount Pictures Corporation; Sony Pictures Entertainment Inc.; Twentieth Century Fox Film Corporation; Universal City Studios LLC; and Warner Bros. Entertainment Inc. MPAA is a proud champion of intellectual property rights, free and fair trade, innovative consumer choices, freedom of expression and the enduring power of movies to enrich and enhance people's lives.