

Website removal from search engines due to copyright violation

Artur Strzelecki

*Department of Informatics, University of Economics in Katowice,
Katowice, Poland*

Received 10 June 2018
Revised 14 August 2018
25 September 2018
Accepted 1 October 2018

Abstract

Purpose – The purpose of this paper is to clarify how many removal requests are made, how often, and who makes these requests, as well as which websites are reported to search engines so they can be removed from the search results.

Design/methodology/approach – Undertakes a deep analysis of more than 3.2bn removed pages from Google's search results requested by reporting organizations from 2011 to 2018 and over 460m removed pages from Bing's search results requested by reporting organizations from 2015 to 2017. The paper focuses on pages that belong to the .pl country coded top-level domain (ccTLD).

Findings – Although the number of requests to remove data from search results has been growing year on year, fewer URLs have been reported in recent years. Some of the requests are, however, unjustified and are rejected by teams representing the search engines. In terms of reporting copyright violations, one company in particular stands out (AudioLock.Net), accounting for 28.1 percent of all reports sent to Google (the top ten companies combined were responsible for 61.3 percent of the total number of reports).

Research limitations/implications – As not every request can be published, the study is based only what is publicly available. Also, the data assigned to Poland is only based on the ccTLD domain name (.pl); other domain extensions for Polish internet users were not considered.

Originality/value – This is first global analysis of data from transparency reports published by search engine companies as prior research has been based on specific notices.

Keywords Search engines, Data retrieval, Search results, Google, American digital millennium copyright act (DCMA), Copyright violation

Paper type Research paper

Introduction

Many technology companies process vast amounts of data. These include online search engines, social networks, software-manufacturing companies, manufacturers of computer and mobile devices and providers of services available on the internet. Depending on the service(s) they provide, these companies may possess data concerning their users or the activities performed while taking advantage of the offered services.

Many of these large technology companies like Google[1], Facebook[2]; Microsoft[3], Twitter[4]; Apple[5]; LinkedIn[6], Snap[7], Pinterest[8], Dropbox[9], Cloudflare[10], Oath[11] publish transparency reports in which they reveal what actions they take concerning the processed data as well as who initiated these actions. Data published in transparency reports may be requested by government agencies from various countries, non-government organizations, companies or private entities.

Government agencies usually require technology companies to share data concerning specific users of the company's services or demand the removal of publicly available data. Data on such requests are usually published every six months and include the number of applications sent from a given country and the type of request, i.e., whether it is a request to share data or remove content. State institutions, courts and parties in civil cases often request users' data from telecommunications and technical companies. National government bodies may ask technology companies to remove access to published content that, in their opinion,

The author would like to thank the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper.



violates state legislation. Such applications are analyzed to determine whether the content actually violates the provisions of national law. If upheld, access to such content in a given country or territory is blocked. Some applications are justified on the grounds of defamation of character or other allegations, including violating local laws that prohibit the incitement of hatred or publishing adult-only contents. In such cases, the law differs in different countries. In their reports, companies present information concerning the number and type of requests they receive from government agencies. They publish the information in order to show the impact of government actions on users and the free flow of information on the internet.

Non-government organizations, companies and private entities usually send requests to remove publicly available content due to copyright violation. Types of content that may be subject to removal depend usually on the technology company and the manner in which it shares the data. In terms of search engines, search results for specific websites may be removed. For social networks, the types of content removed include individual posts, photos, videos, advertisements, profiles, accounts, sites, groups and events. For data-storage and data-sharing services, the types of content removed include stored text documents and graphic, audio or video files. A request sent by a copyright owner is usually processed by a team responsible for intellectual property and copyright. If the team determines that the application is complete and correct then the requested content is removed.

Private entities residing within the European Union and the European Economic Area may request the removal of data from a search engine in order to protect personal data. In May 2014 the Court of Justice of the European Union, when considering the case of a resident of Spain (Frantziou, 2014), Mario Costeja González, against Google, determined that every private entity has the right to request that search engine operators such as Google or Bing remove search results including the name and surname of such an entity. The search engine's operator must comply with such a request if the links indicated in it lead to information which is improper, exaggerated, inadequate, or insignificant, taking into consideration the public interest, including such factors as the role of that person in public life. The links are removed only from results pertinent to the request, including the name and surname of that person. Concerning searches within the European Economic Area, URLs are removed from all European search results. With the use of geolocation, the access to URLs is restricted based on the country of the person requesting their removal. If a query in the form of a name and surname is entered into the search engine, then the following message is displayed below the search results: "Some results may have been removed under data protection law in Europe."

Technology companies usually have forms available for requesting the removal of content. The forms are sent by government agencies, copyright owners and private entities that want to take advantage of the right to be forgotten. The form usually requires a declaration that the requestor is the owner of copyright or content or is the person whose data is being processed, as well details of from where the content should be removed, or stopped from being displayed in the case of search results. A request to remove content based on copyright violation should only be processed after receiving an official request form; however, in order to take advantage of the right to be forgotten, one has only to confirm one's identity with an appropriate identity document.

The rest of this paper is organized as follows. In the next section, a review of the relevant literature is undertaken. Then the context of study and study setup is described and, following this, the results are presented and discussed. Finally, conclusions are drawn.

Literature review

An analysis of the literature regarding search engines reveals two main fields. The first consists of research conducted to understand how search results are generated and what impact they have, including both organic results and sponsored search results. The second concerns reports of copyright violations and the proposed solutions.

Search engine results pages

Search results are divided into two main categories: organic and sponsored (Jansen and Resnick, 2006). Initially, organic results were tested in terms of how they were perceived and utilized by the users. Much research (Granka *et al.*, 2004; Hotchkiss *et al.*, 2005) has confirmed that, for organic searches, users most often look at the first and then the second result. Results for “clickability” were similar. Users not only looked at the highest results the most, but also clicked on them most often (Pass *et al.*, 2006; Zhang and Moffat, 2006; Chitika Insights, 2013; Font-Julian *et al.*, 2018).

There has also been research concerning to what extent users are able to identify which results are organic and which ones are sponsored (Lewandowski *et al.*, 2018). Many of the undertaken academic works have discussed which elements have the strongest impact on the order of organic results. These include the analysis of links (Bifet *et al.*, 2005; Fortunato *et al.*, 2006), contents (Evans, 2007; Moreno and Martinez, 2013) and IP addresses (Strzelecki, 2017). Other research has analyzed what, and how often, users search (Andersson, 2017; Dotson *et al.*, 2017).

Much research has focused on the issue of what strategies to undertake for sponsored results in order to achieve optimal effects (Sen, 2005; Cheng *et al.*, 2018). Users prefer also results to be more adjusted to their expectations, or more customized, and eventually accept the changes to the displayed results (Zhitomirsky-Geffet *et al.*, 2016). Users also tend use the search engine more when they see an interesting advertisement on the TV (Mukherjee and Jansen, 2017).

Copyright violations and search engines

Copyright can be infringed in various media formats. The Intellectual Property Office (Kantar Media, 2018) reported that, in the UK, it is estimated that 15 percent of UK internet users aged 12 and over consumed at least one item of online content illegally in the past three months. This estimate is based on a combination of six key content types (music, film, TV programs, computer software, e-books and video games).

The YouTube service, which is owned by Google, has been the subject of many copyright violation reports (Bridy, 2016; Sag, 2017; Jacques *et al.*, 2017). However, the mechanisms of copyright protection introduced are different from the ones used in the Google search engine (Brown, 2008).

Copyrighted content is removed from YouTube. Erickson and Kretschmer (2018) studied take-downs within an original data set of 1,839 YouTube music-video parodies and found an overall rate of take-downs within the sample of 32.9 percent across the four-year period. Other research has revealed that YouTube is the most vulnerable platform for notices of copyright infringement (Gray, 2015). Heald (2015) used the transaction cost theory to track 90 songs on YouTube that reached No. 1 on the USA, French and Brazilian pop charts from 1930 to 1960: the data collected included the identity of the uploader, type of upload, number of views, date of upload and monetization status.

One of the studies in this area considering the legal perspective analyzed 816 requests sent in 2006 (Urban and Quilter, 2006). Although research in this area remains limited, Zhang *et al.* (2017) have recently proposed a model of the criteria for achieving more accurate DCMA take-down notices. Google processes DMCA requests for websites that are not even included in the search index (Donaldson, 2017). If the request proves to be valid, then the URLs will never appear in search results; therefore, the number of requests received by Google does not represent the actual number of links presented in search results.

Reports and requests based on the DMCA may be used to manipulate the search results and remove unfavorable opinions (Urist, 2006). The Google search engine takes into consideration the opinions of copyright owners and introduces changes in its algorithm, including changing the functioning of the auto-filling mechanism (Bridy, 2016).

DMCA reports and requests can also be used to remove graphics from search results (Sag, 2017) as well as check what advertisements are displayed to users on websites that have already been reported to the search engine and are included in Lumen (Watters *et al.*, 2014). In the vast majority of cases, these are high-risk advertisements (Watters, 2013), i.e., those including malware, gambling, false software, websites installing software, etc. Websites present in Lumen, previously reported to the search engine, usually do not include an age warning. Content that appears there may not only violate copyrights but may also be inappropriate for minors (Watters, 2014).

The Google News service has been highlighted as violating copyrights (Allgrove and Ganley, 2007). This automatic software copies fragments of contents from press portals and publishes them on the Google News (Calzada and Gil, 2017).

The Google Cache service, however, i.e., the last copy of a website which that has already been indexed in the search engine, is not regarded as a violation of copyright (Peguera, 2008). However, many automatic functions of the search engine have been reported as a copyright violation (Fitzgerald *et al.*, 2008).

Context of study

Search engines regularly receive requests to remove content from search results that may violate the law. Published reports include information concerning requests to remove links from search results and are intended to make it easier for all users to understand how copyright impacts the availability of content in online search engines. The Bing search engine publishes general reports providing only four pieces of information regarding removing search results following a report concerning copyright violation: the number of requests sent to remove search results; the number of reported URLs; the number of removed URLs; and the number of kept URLs. The Google search engine publishes complete information concerning all sent reports and requests to remove URLs from search results, as well as the undertaken actions.

The Google search engine based the mechanism for removing results on the American Digital Millennium Copyright Act (DCMA). The DMCA is a copyright act in force since 1998 in the USA, prohibiting the creating and distributing of technology with the use of which it is possible to violate the digital mechanisms of copying restrictions (www.copyright.gov/legislation/dmca.pdf).

According to Google's terms and conditions, it addresses clearly formulated, specific requests concerning a supposed violation of copyright. The online form provided is in accordance with the DMCA and provides a mechanism for copyright owners from all over the world to initiate the process of removing content from search results if they believe that a URL leads to content violating copyrights. After receiving a valid request to remove data, a team representing the search engine verifies whether it is complete and meets all of the requirements. If the request is complete and justified, the URL is removed from search results. Google accepts also reports from trusted partners that have been invited to the Trusted Copyright Removal Program. This program currently includes 114 (<https://blog.google/topics/public-policy/continuing-to-create-value-while/>) partners and they can send large volumes of reported URLs.

The Bing search engine also accepts requests concerning the violation of copyrights in its search results based on the DMCA. Bing accepts reports via three communication channels: e-mail; form; and APIs for selected reporting agencies. The third channel is reserved for partners sending many URLs.

Other search engines like the Chinese Baidu, Shenma, Haosou, and Sogou, the Russian Yandex and Mail.ru, the Japanese Yahoo, the Czech Seznam, the Vietnamese CocCoc and the US DuckDuckGo do not share any data about website removals. Only the Korean Naver shares some basic data about requests made by governments, but not those made by

copyright owners. Yahoo.com, AOL and MSN all use Bing's search engine and data for these companies is included in Bing's report.

The content of all requests sent to Google and Bing is available in the Lumen database (previously known as the Chilling Effects Clearinghouse). The Lumen database (www.lumendatabase.org/) constitutes a project of the Berkman Klein Center for Internet & Society institute at Harvard University. Lumen works with many different partners in terms of international research and provides information about the global situation in terms of requests to remove content from the internet. The Lumen database collects and analyzes legal complaints and requests for website removal, helping internet users to know their rights and understand the law. These data enable the study of the prevalence of legal threats and let internet users see the source of content removals. Lumen publishes and analyzes various types of requests to remove data from the internet, including requests based on copyright. Lumen receives information concerning such requests not only from companies but also private entities. When it is allowed by the law, Google displays links to requests published by Lumen instead of the removed content.

The goal of the present study is to find the answers to the following research questions (RQs):

- RQ1.* How many requests were sent to online search engines and how many URLs have been removed?
 - RQ1a.* What is the share of requests to remove content based on copyright infringement in the Google search engine for Polish websites?
 - RQ1b.* Which Polish websites are most often indicated as violating copyright and how often are these websites' URLs removed?
- RQ2.* What knowledge may be discovered by fully analyzing data concerning requests to remove content from various time periods?
- RQ3.* How often are copyrights violated and which (type of) copyright holders are concerned?

Data and methodology

The data set for the Bing search engine was downloaded on May 15, 2018 from the website on which Bing publishes its transparency reports. The amount of published data regarding removing content from search results due to copyright violation was small when compared to the amount of data published by Google. Data from Bing allows only for a general comparison with data coming from Google. Bing shares six files with an .xlsx extension that include the basic information covering a six-month period. The data available at the time this research was undertaken covered the period from January 1, 2015 to December 31, 2017. The .xlsx files included information concerning requests to remove content in four categories: requests from governments such as claims of violations of local laws; requests from European and Russian residents to filter search results about them for queries that include their names; requests from copyright owners claiming infringement of protected works; and requests from individuals to remove nonconsensual pornography, which is the sharing of nude or sexually explicit photos or videos online without consent.

The data set for the Google search engine was downloaded on April 14, 2018 from the website on which Google publishes its transparency reports. The size of the compressed archive was 3GB. After extracting, the size of the data included in three files (requests.csv, domains.csv and urls-no-action-taken.csv) was 20GB. The information was presented in text form. The requests.csv file included 594,1076 lines of data, the domains.csv included 206,181,694 lines of data and urls-no-action-taken.csv included 123,437,248 lines of data. The fields in .csv files were divided by commas and some text values, which included a comma in

the string of characters, were embedded in quotation marks. Apart from three files with data, the archive included a description of the files' contents.

The `requests.csv` file contained the information for all requests for URLs to be delisted from web search results for copyright violations. The information was organized by the unique request ID for each copyright removal request, the date (in UTC) that the request was received (in ISO 8601 format), the URL to the Lumen page documenting the request, the ID number of a unique copyright owner, the name of the copyright owner associated with the request, the ID number of the unique reporting organization, the name of the reporting organization associated with the request, the number of URLs that were specified in the request that were removed, the number of URLs that were specified in the request but not removed, the number of URLs that were specified in the request that were still pending review.

The `domains.csv` file contained the information for all domains that had a URL requested to be delisted from web search results for copyright violations. The information was organized by the unique request ID for each copyright removal request, the normalized domain specified within the request, the number of URLs that were specified in the request that were removed, the number of URLs that were specified in the request but not removed, the number of URLs that were specified in the request that were still pending review.

The `urls-no-action-taken.csv` file contained the information for all URLs requested to be delisted from web search results for copyright violations for which moderators took no action on the request. The information was organized by the unique request ID for each copyright removal request, the normalized domain specified within the request, a URL that was specified in the request but not removed.

Data from the `domains.csv` and `requests.csv` files were loaded on to the MySQL database. The prepared database instance was based on the MySQL Community Server 5.7, MySQL Workbench to execute SQL queries and reading the results and the computer on which the database was installed was a MacBook Pro with MacOS High Sierra 10.13.4, RAM 8 GB 1867 MHz DDR3, a 120 GB SSD and a 2.7 GHz Intel Core i5.

Table structures representing the data included in the `.csv` files were prepared in the database. Then these structures were filled with data, taking into consideration that the fields were divided with a comma and some text values were embedded in quotation marks. In the `requests` table, the first-time marker of a sent request constituted the day March 10, 2011.

During the first stage of calculations, all data from the `domains.csv` file was included in the `domains` table. During the second stage of calculations, the table was loaded only with data concerning domains. These were domains marked as country coded in Poland. The reason why Polish websites were chosen for detailed analysis is that, in top six domains mentioned in Google's transparency report (<https://transparencyreport.google.com/>), five of them are generic top-level domains like `.net` (two domains), `.com`, `.xyz`, and `.co`, while the sixth domain is `.pl`, i.e., a Polish domain space.

Extraction of lines only including `.pl` domains was executed with the use of a terminal's `grep` command. The `grep` command in this case was used to search and divide lines including strings of characters in the text which fit the provided regular expression. `Grep` finds all strings which include the `".pl,"` fragment, clearly identifying lines including data concerning domains marked as Polish country-coded domains. The data entered during the second stage of calculations into the `domains` table were limited only to `.pl` domains in order to speed up calculations. All data included in the `requests.csv` file was loaded into the `requests` table, with researches and calculations being carried out on their base. In both tables, key columns received text indexes in order to speed up the calculations. In the `domains` table, the indexes were assumed for columns including request id and normalized domain. In the `requests` table, indexes were assumed for columns id, data, id number of a unique copyright owner and the id number of the unique reporting organization.

Results

Basic information concerning data from Google's transparency report is presented in Table I. The first column includes data concerning all records and the second column only .pl domains.

In Table I the number of notices constitutes the total number of all reports sent to the search engine with a request to remove content due to copyright violation. The date of the earliest entry was March 10, 2011 and the record also included the lowest ID in the entire data collection. "Number of domains" constitutes the total number of domains of the highest level (for example google.com), which includes URLs of content that has been requested to be removed. "Number of URLs" constitutes the total number of all URLs requested to be removed. "Copyright owners" constitutes the total number of separate people or entities that have stated that they have the sole right to the content requested to be removed due to copyright violation. "Reporting organization" stands for the total number of separate entities or organizations addressing Google on behalf of copyright owners to remove search results due to copyright claims. "URLs removed" stands for the total number of URLs that have been removed from the search results due to a sent request. "URLs no action" refers to the number of URLs which were not removed from search results.

The analysis of requests to remove content from PL domains covered only content included on Polish websites. However, it has to be remembered that there is the possibility that websites may be in Polish but do not have to end with a .pl domain. In this case, these are usually functional domains such as .com or .net.

Analyzing only a part of the data based on .pl domains, illustrated in Table I, the following results are produced. Two characteristics significantly stand out from the data. The first characteristic is that 12.8 percent of reports included a request to remove content from .pl domains. This means that every eighth request concerned copyright violation on a .pl domain. The second characteristic is that 8.8 percent of copyright owners requested the removal of content from .pl domains. Requests to remove URLs with a .pl domain from search results were made by 3.3 percent of the total number of organizations representing copyright owners. The remaining calculated values show that the 13,035 reported domains constituted 0.7 percent of all reported domains, and the number of URLs requested to be removed constituted 1.1 percent of all reported URLs. Regarding undertaken actions, 1.1 percent of URLs removed from search results were on .pl domains, and 0.8 percent of URLs not removed were on .pl domains.

From an analysis of Table II, concerning all requests over the specified period, it is evident that, in each period, there was an increase in the number of requests sent concerning the removal of content. In 2011, the data concerning removing content was just starting to be published and it is the year in which the number of reports was the smallest. In 2012 the total number of requests was 427,979 or 1,170 requests to remove content each day. Treating this value as a starting point, in 2013, it remained at a similar level as there were 1,315 requests

| | Overall | For .pl | Percentage |
|------------------------|---------------|------------|------------|
| Number of notices | 5,940,364 | 763,119 | 12.8 |
| Number of domains | 1,694,890 | 13,035 | 0.8 |
| Number of URLs | 3,278,140,393 | 38,598,269 | 1.2 |
| Copyright owners | 136,117 | 12,038 | 8.8 |
| Reporting organization | 121,414 | 4,021 | 3.3 |
| URLs removed | 3,079,881,303 | 36,193,566 | 1.2 |
| URLs no action | 298,259,067 | 2,404,703 | 0.8 |

Table I.
Summary of web
search copyright
removals from Google

Source: <https://transparencyreport.google.com/>

Table II.
Number of request
sent to Google

| Year | Number of requests | Requests of .pl | Percentage |
|-----------------------|--------------------|-----------------|------------|
| 2011 | 35,487 | 3,071 | 8.7 |
| 2012 | 427,979 | 35,905 | 8.4 |
| 2013 | 480,012 | 60,775 | 12.7 |
| 2014 | 831,080 | 108,908 | 13.1 |
| 2015 | 1,190,486 | 139,706 | 11.7 |
| 2016 | 1,251,827 | 179,170 | 14.3 |
| 2017 | 1,298,834 | 189,391 | 14.6 |
| 2018 – do 13 kwietnia | 424,659 | 46,193 | 10.9 |

Source: <https://transparencyreport.google.com/>

sent each day on average, i.e., an increase of 12.4 percent. However, in the two following years a constant and equal increase in the number of reports was observed. In 2014, each day there was an average of 2,277 reports, which constitutes an increase at 73.1 percent, and in 2015 there were already 3,262 reports sent each day, i.e., an increase of 43.2 percent. In the following years, this number grew slower than before, with the figure reaching 3,420 in 2016 (an increase at 4.6 percent) and 3,558 requests each day in 2017 (an increase of 4.0 percent). In 2018, until the day of downloading the data, an average of 4,122 requests a day had been sent (an increase of 13.7 percent), i.e., higher than during the two previous years.

Analyzing only a part of the data based on .pl domains, as illustrated in Table II, the following results are revealed. There were not many requests in 2011, when the reports were initially made public but, in the following years, an increase in the number of reports can be seen, although the increase is not uniform. In 2012 and 2013 the total number of reports rose steadily, with reports for the .pl domain increasing by almost 70 percent each year. In 2014, however, the number of reports increased by 44.2 percent, a much smaller increase compared to the global trend. In 2015, the increase for .pl domains was smaller again at 28.2 percent. However, in 2016 the trend clearly changed and the increase of 28.2 percent was the highest for all reports globally. In 2017, the increase was 5.7 percent, similar to the global increase.

When comparing the total changes with changes in terms of just one domain it is possible to draw conclusions that copyright violations in specific countries differ in their frequency and intensity. The analyzed section of data shows that each year there are more violations but the numbers grow differently in different countries.

Information on the number of reports sent shows a high irregularity in these amounts (see Figure 1). As shown in Table II, there were more reports sent each year but the increase in the number of reports was not linear. The horizontal axis in Figure 1 presents the date of sent reports, while the vertical axis shows how many reports were sent on that day. The chart shows also the line of the trend, which was calculated as an average from 30 values, and shows that the frequency of taking advantage of such services is very diverse. Apart from a constant annual increase, it is impossible to find a regularity or trend that would predict the figures for the following years as the distribution of the number of sent requests is different. The first part of 2012 had the smallest number of reports sent, on average, while the second part of the year had much more than the average value observed in the following year. Since 2015, it is possible to observe that the trend line increases and drops irregularly, reflecting periods when there were many more reports being sent than in others.

Table III presents results comparing the number of removed URLs with the number of URLs kept in search results on Google. During the entire researched period the number of removed URLs was 91.1 percent of all URLs requested to be removed. During the entire

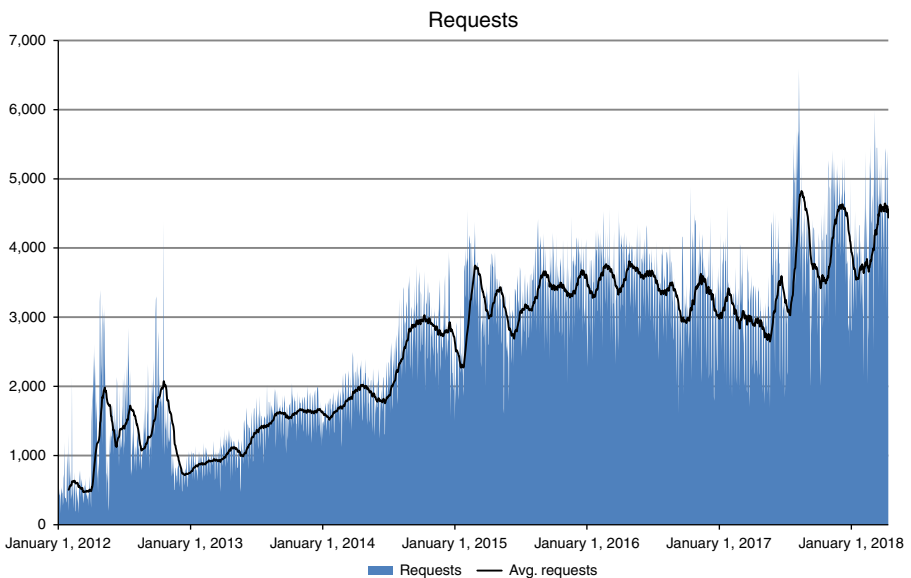


Figure 1.
Requests number over
time (2012–2018)

| Year | URLs of removed | URLs of rejected | % of removed | % of rejected |
|-----------------------|-----------------|------------------|--------------|---------------|
| 2011 | 3,197,883 | 269,487 | 92.2 | 7.8 |
| 2012 | 54,334,134 | 7,012,798 | 88.5 | 11.5 |
| 2013 | 221,954,446 | 21,436,889 | 91.1 | 8.9 |
| 2014 | 321,713,821 | 25,634,003 | 92.6 | 7.4 |
| 2015 | 519,974,266 | 41,054,927 | 92.6 | 7.4 |
| 2016 | 914,794,045 | 103,389,231 | 89.8 | 10.2 |
| 2017 | 827,276,921 | 89,450,051 | 90.2 | 9.8 |
| 2018 – do 13 kwietnia | 216,635,787 | 10,011,681 | 95.6 | 4.6 |

Table III.
Number of removed
and rejected URLs
from Google

Source: <https://transparencyreport.google.com/>

researched period, the share of URLs removed from Google's index changed, being either higher or lower than in following years. Table III also shows that the most URLs were removed in 2016 (more than 1bn URLs). This also shows the amount of work done by the teams representing the search engines in accepting or rejecting requests to remove URLs from search results. In 2016, the most rejections concerning removing reported URLs (more than 103m URLs). As previously highlighted in Table II, the number of reports increased each year, but the number of URLs requested to be removed decreased because these reports included fewer search results for removal. If the decreasing trend continues in the second half of 2018, this will mean that there will be even fewer URLs reported in terms of copyright infringements.

In comparison, the Bing search engine, in the second half of 2017, removed 99.90 percent of all reported URLs. It must be noted that the total number of URLs sent to Bing was much smaller than the number sent to Google. Table IV presents data published by Bing concerning the number of reports sent and URLs accepted and rejected. Available data from 2015 and 2016 show that the number of reports sent has constantly increased. In 2015 there were 1,996,276 reports while, in 2016, the number grew to 6,677,719. In the first half of 2017,

approximately 16m requests were sent, and in the second half of 2017, more than 19m requests were sent.

There is no data explaining the reasons behind such a difference in the number of requests sent to Google and Bing. One possible reason is that the form for sending reports to Bing is more useful than its Google equivalent, although that should be researched separately. The second reason may result from the fact reports sent to Google concerning a given internet domain are treated as a negative ranking signal for that domain. Thus, the search engine does not display much content from domains that have been previously reported. Therefore, the number of reports sent does not increase as quickly for Google as it does for Bing, which perhaps does not consider this as a ranking signal and thus the number constantly grows. However, these are just hypotheses that need further research.

Analyzing only a part of the data based on .pl domains (see Table V), the following results are revealed. During the entire researched period the number of URLs removed from .pl domains was 93.8 percent of the total number of URLs requested for removal. For .pl domains, this constitutes a larger share of removed URLs than for all removed URLs globally. Until 2015 the number of reported URLs had grown each year but subsequently decreased. This is perhaps the result of an improvement in terms of respecting copyrights, thus publishing less content violating these rights. What is interesting is that the largest number of removed URLs globally, which took place in 2016, does not correlate with the largest number of reports in the .pl domain. It should also be noted that, after 2015, the number of removed URLs remained above 95 percent of the total number of URLs reported in all domains. This might mean that the quality of reports, including URLs from the .pl domain, was high and that the requests, therefore, included fewer URLs that were rejected.

In terms of the top-10 list of .pl domains with the highest number of reports (see Table VI), the chomikuj.pl domain stands out. Chomikuj.pl is the fifth most reported domain in the entire data set shared by Google. When analyzing the number of reports, one has to remember that in one report it is possible to include more than one domain. The total number of reports including URLs from the .pl domain was 763,119 and, as can be seen when compared to other specific domains, the chomikuj.pl is listed in almost half of the reports. chomikuj.pl includes an online service that allows users to store files and share them publicly or with other users. Users often take advantage of chomikuj.pl to share copyright-protected materials, explaining the large number of reports. Other domains in the top-10 list also included services that allow the sharing and downloading of files. These services allowed copyright-protected materials to be found and downloaded using of a torrent search engines or links to websites that stored the copyrighted materials.

The top-10 list of .pl domains with the largest number of removed URLs (see Table VII), also included the chomikuj.pl domain, ranked highest. URLs removed from this domain constituted 62 percent of all URLs removed from the .pl domain. The list also included also two domains, freed0m4all.pl and zajumaj.pl, that had relatively much fewer reports than

| | Requests | URLs requested | URLs accepted | URLs rejected | % of URLs accepted |
|---------|------------|----------------|---------------|---------------|--------------------|
| 2015 H1 | 1,020,142 | 24,520,508 | 22,462,834 | 2,057,672 | 92 |
| 2015 H2 | 976,134 | 59,473,002 | 58,487,912 | 985,090 | 98 |
| 2016 H1 | 2,548,451 | 91,781,926 | 91,269,366 | 512,560 | 99 |
| 2016 H2 | 4,129,268 | 165,601,360 | 165,285,689 | 315,671 | 99.81 |
| 2017 H1 | 16,268,707 | 121,541,381 | 121,111,170 | 430,211 | 99.65 |
| 2017 H2 | 19,151,385 | 123,703,166 | 127,570,775 | 132,391 | 99.90 |

Source: www.microsoft.com/en-us/about/corporate-responsibility/crrr

Table IV.
Number of removed
and rejected URLs
from Bing

others but many URLs from these domains were removed from search engine results. Also, in terms of these two domains, the number of URLs that were not removed was relatively low, which means that almost all the reported content violated copyright according to Google's checking procedures. The analysis of specific fields such as reports, removed URLs and kept URLs, shows how important it is to granulate data and analyze it both in terms of specific fields as well as on the level of individual domains. The information gathered shows the different shares various domains have in publishing content violating copyright and how different the frequency is of removing this content from search results.

Table V.
Number of removed and rejected URLs in the .pl domain

| Year | URLs removed | URLs rejected | % of removed | % of rejected |
|-------------------|--------------|---------------|--------------|---------------|
| 2011 | 23,802 | 2,625 | 90.0 | 10.0 |
| 2012 | 647,884 | 63,646 | 91.0 | 9.0 |
| 2013 | 1,504,997 | 203,101 | 88.1 | 11.9 |
| 2014 | 3,598,752 | 230,410 | 93.9 | 6.1 |
| 2015 | 11,111,572 | 928,102 | 92.2 | 7.8 |
| 2016 | 9,958,423 | 517,769 | 95.0 | 5.0 |
| 2017 | 8,176,589 | 403,546 | 95.2 | 4.8 |
| 2018 – 13th April | 1,171,547 | 55,504 | 95.4 | 4.6 |

Source: <https://transparencyreport.google.com/>

Table VI.
Top-10 .pl domains in terms of number of requests for removal

| Domain name | Number of requests | Number of removed | Number of rejected |
|----------------|--------------------|-------------------|--------------------|
| chomikuj.pl | 321,886 | 22,437,686 | 1,174,145 |
| ulub.pl | 68,212 | 556,133 | 13,857 |
| freedisc.pl | 54,093 | 335,687 | 35,257 |
| mp3s.pl | 53,228 | 1,514,242 | 253,785 |
| pobieramy24.pl | 51,029 | 391,587 | 69,421 |
| darkwarez.pl | 42,928 | 250,712 | 22,952 |
| exsite.pl | 34,311 | 160,979 | 22,763 |
| katproxy.pl | 27,237 | 162,017 | 3,694 |
| torrenty.pl | 26,423 | 100,360 | 7,954 |
| wrzuta.pl | 26,400 | 324,182 | 45,988 |

Source: <https://transparencyreport.google.com/>

Table VII.
Top-10 .pl domains in terms of number of removed URLs

| Domain name | Number of requests | Number of removed | Number of rejected |
|----------------|--------------------|-------------------|--------------------|
| chomikuj.pl | 321,886 | 22,437,686 | 1,174,145 |
| fileshark.pl | 26,213 | 3,571,260 | 113,017 |
| mp3s.pl | 53,228 | 1,514,242 | 253,785 |
| ulub.pl | 68,212 | 556,133 | 13,857 |
| freed0m4all.pl | 1,411 | 504,943 | 4,022 |
| unblocked.pl | 10,791 | 447,896 | 13,280 |
| pobieramy24.pl | 51,029 | 391,587 | 69,421 |
| zajumaj.pl | 4,821 | 342,225 | 470 |
| freedisc.pl | 54,093 | 335,687 | 35,257 |
| wrzuta.pl | 26,400 | 324,182 | 45,988 |

Source: <https://transparencyreport.google.com/>

The top-10 list of .pl domains with the largest number of URLs not removed (see Table VIII) also includes the chomikuj.pl domain in first position. This shows that some of the content reported and requested to be removed in reality did not violate copyright and that reporting agencies may have been abusing their rights and reporting URLs that did not violate copyright. This can be seen especially when analyzing the reports for the tekstowo.pl domain. This domain was reported 4,428 times as including content violating copyright. However, only 29 URLs were removed from search results, and more than 210,000 URLs were not removed. This domain publishes song lyrics and translations. Organizations reporting copyright violations had searched for materials protected by copyright using the name of the entire work. o. The tekstowo.pl domain only publishes song lyrics and does not host audio or video formats. Lyrics are also copyrighted; however, it is not known why these requests were rejected. That topic needs further research.

Owners of copyrights usually constitute usually commercial organizations that handle publishing works protected by copyrights (Table IX shows the top-10 copyright owners in terms of number of requests but does not include individual authors). They usually transfer a part of their rights or grant a license for the use their works by commercial organizations that

| Domain name | Number of requests | Number of removed | Number of rejected |
|---------------------|--------------------|-------------------|--------------------|
| chomikuj.pl | 321,886 | 22,437,686 | 1,174,145 |
| mp3s.pl | 53,228 | 1,514,242 | 253,785 |
| tekstowo.pl | 4,428 | 29 | 210,855 |
| fileshark.pl | 26,213 | 3,571,260 | 113,017 |
| pobieramy24.pl | 51,029 | 391,587 | 69,421 |
| wrzuta.pl | 26,400 | 324,182 | 45,988 |
| thepiratebay.net.pl | 3,688 | 171,921 | 36,336 |
| freedisc.pl | 54,093 | 335,687 | 35,257 |
| darkwarez.pl | 42,928 | 250,712 | 22,952 |
| exsite.pl | 34,311 | 160,979 | 22,763 |

Source: <https://transparencyreport.google.com/>

Table VIII.
Top-10 .pl domains in
terms of number of
URLs not removed

| Number of requests | Owner name | Registration country |
|--------------------|--|----------------------|
| 232,178 | BPI Ltd Member Companies | United Kingdom |
| 142,310 | BPI (British Recorded Music Industry) Ltd | United Kingdom |
| 79,137 | Nuclear Blast Records | Germany |
| 48,140 | IFPI | United Kingdom |
| 44,402 | Universal Music GmbH | Germany |
| 41,734 | Beggars Group Digital Ltd | United Kingdom |
| 36,166 | Warner Music Group Germany Holding GmbH | Germany |
| 35,758 | Sony Music Entertainment Germany GmbH | Germany |
| 35,295 | Entertainment One | Canada |
| 34,800 | RIAA member companies (EMI Music North America, Sony Music Entertainment, Universal Music Group, Warner Music Group) | United States |

Source: <https://transparencyreport.google.com/>

Table IX.
Top-10 copyright
owners in terms of
number of requests

publish them. Among the organizations whose rights were most often violated, the main offices of the top-eight companies were located in the UK and Germany, although these organizations usually represent people from all over the world. The two remaining organizations were registered in North America and represented mainly the interests of Americans and Canadians.

Although organizations that possess copyrights may report copyright violations to search engines themselves, the largest organizations commission other companies to do it, i.e., companies that specialize in searching for content violating copyright and reporting this to search engines.

Organizations representing copyright owners which (see Table X) are responsible for most requests for Google to remove URLs. The most active organization was AudioLock.Net, which had made 1,673,730 copyright violation reports. It constituted 28.1 percent of all reports sent to Google. The entire top ten was responsible for 61.3 percent of the total number of reports.

When viewing individual reports sent to the search engine, it is notable that the largest ones, in terms of the number of URLs sent in one report, included more than 32,000 URLs. These reports were almost entirely accepted by the search engine. However, there were also reports including more than 28,000 URLs that were completely rejected. An analysis of the detailed reports shows that each report is analyzed separately by the team representing the search engine and that all URLs included in the report were verified.

Discussion

The conducted analysis shows that online search engines are engaged in the process of copyright protection. In cases confirming the validity of a request in accordance with the copyright protection law, the content is removed from the services offered by the search engine, i.e., the URL is removed from search results. After removing the content, if the search engine is able to inform the owner of the website then it does so (for Google, via the Google Search Console; for Bing, via Bing Webmaster Tools).

In accordance with the guidelines followed by the search engines, they respond to clearly formulated, specific requests concerning a supposed violation of copyrights. It may happen that, after verifying the request, it is determined that some URLs reported clearly do not violate copyrights. In such cases, the request is rejected and the URL is not removed from the search engine's results. A rejection may also occur if there is insufficient information concerning why a given URL violates copyright, if content included in the report was not found, or if it is determined that the content is being used correctly. Search engines also sometimes receive imprecise or unjustified requests to remove content from search results, referring to content that does not violate copyright; these requests are also rejected.

When the website owner receives a notice of DMCA removal from search engine and feels that the sites or pages were mistakenly removed due to a DMCA request, a counter

| Number of requests | Reporting organization name |
|--------------------|---|
| 1,673,730 | AudioLock.NET |
| 429,012 | MUSO.com Anti-piracy |
| 411,103 | BPI (British Recorded Music Industry) Ltd |
| 311,319 | proMedia |
| 238,320 | Digimarc |
| 171,641 | Counterfeit.Technology |
| 164,408 | Link-Busters.com |
| 85,210 | rivendell |
| 78,592 | Total Wipes Music Group |
| 78,104 | Topple Track |

Source: <https://transparencyreport.google.com/>

Table X.
Top-10 organizations
in terms of number
of requests made
to Google

notification can be submitted. Then, search engine can reinstate these materials into search results upon the receipt of a DMCA counter notification. This shows that it can happen that legitimate content or content with fair use are sometimes being removed from search engine. This solution is far from perfect, however, it is based on the DMCA and search engines are obliged to follow this act. If search engine would not obey this regulation, it could be subject to a claim of copyright infringement, regardless of its merits.

As Zhang *et al.* (2017) observed that the accuracy of notices is not known, and search engines do not provide clear information about how they check the legitimacy of these requests. It is not known if and how search engines check the legitimacy of allegedly infringing content or what criteria they use for these actions. This study shows that the practice of removing data from search engines is still rising and the reports are very transparent. However, the exact rules and criteria of removing content are not disclosed.

Conclusion

The main conclusions can be drawn from of the present study, related to the three RQs. *RQ1* was answered by results confirming the fact that search engines have allowed requests to remove contents for more than eight years now (since 2011). These requests have resulted in Google removing more than 3 billion URLs (within eight years) and Bing removing more than 590 million URLs (within three years). A total of 91.1 percent of all requested URLs were removed from Google and 99.2 percent from Bing.

An analysis of a portion of data based solely on URLs in the .pl domain revealed that the fields concerning one selected domain may have a different share in reports and removals and that this share is not proportional. Reports concerning the .pl domain stand out with an over-proportional share of the number of reports sent and an over-proportional number of copyright owners which these reports concerned when compared to the number of URLs removed. Requests for the .pl domain stand out also with a higher percentage of accepted reports. In the .pl domain, 93.8 percent of reported URLs were removed from the Google search engine, but in the last three years the number is even higher (above 95 percent).

Data analysis of .pl domains, in terms of those most often included in the reports and those with the most URLs either removed or kept, showed that one domain had a significant share (chomikuj.pl), accounting for almost half of the reports sent, 62 percent of requests to remove URLs and 48.8 percent of URLs not removed.

Regarding *RQ2*, analyzing the data have shown that the global frequency and increase of removals during the researched period was not identical with the trend visible in the .pl domain. While both show a continual increase in the number of reports, the pace of the growth is different when comparing fragments of data with the entire set.

The answer to *RQ3* is that, although copyright owners have the ability to make requests to search engines, they tend to use services provided by entities that specialize in searching and reporting violations based of the DMCA. The largest such partners of search engines can take advantage of APIs to report large volumes of URLs. Smaller organizations, however, use dedicated forms. Results also revealed that most copyright owners have their main offices in the USA, the UK and Germany.

Removing websites from search engine results pages has an impact on users: when users search for information, removed websites do not appear in the results. If Google has removed any results from its results page, in response to multiple complaints received under the DMCA, Google provides information about this. Users may read the DMCA complaint(s) that caused the removal via direct link to Lumen. Bing, however, informs users that some results have been removed and directs user to a webpage explaining how Bing delivers search results (there is also dedicated section for copyright infringement on this page); however, there is no link to the complaint at Lumen, so users do not know what was removed from Bing's search results.

The majority of removal requests come from music and movie companies. These industries benefit from selling rights for playing and streaming movies, TV series, songs and other creative work in audio or video formats. Companies in these industries may either have licensing agreements with copyright owners or produce their own creative work. If their copyrighted work is available on an illegal website and accessible through a search engine, users may not pay for this content from the legitimate source but instead use source that violates copyright, which explains why music and movie companies send large amounts of removal requests to search engines, i.e., to protect their incomes.

This papers' contribution to the literature is that this is the first comprehensive study regarding website removals from search engines. Previous research has focused only on certain requests and not encompassed them all. One direction for further research is to check whether large amounts of removal requests for certain domains cause a significant demotion in search-engine ranking. This requires a comparison of the data published in the transparency reports with the data available in the services that monitor the visibility of websites in search engines. Another direction for further research is an analysis of, and comparison between, removal requests made for other country domains.

Notes

1. <https://transparencyreport.google.com/>
2. <https://transparency.facebook.com/>
3. <https://microsoft.com/en-us/about/corporate-responsibility/crrr>
4. <https://transparency.twitter.com/>
5. <https://apple.com/privacy/transparency-reports/>
6. <https://linkedin.com/legal/transparency>
7. <https://snap.com/en-US/privacy/transparency/>
8. <https://help.pinterest.com/en/articles/transparency-report>
9. <https://dropbox.com/transparency/reports>
10. <https://cloudflare.com/transparency/>
11. <https://transparency.oath.com/>

References

- Allgrove, B. and Ganley, P. (2007), "Search engines, data aggregators and UK copyright law: a proposal", working paper, Baker & McKenzie LLP, London, available at: <http://dx.doi.org/10.2139/ssrn.961797>
- Andersson, C. (2017), "Google is not fun: an investigation of how Swedish teenagers frame online searching", *Journal of Documentation*, Vol. 73 No. 6, pp. 1244-1260, available at: <https://doi.org/10.1108/JD-03-2017-0048>
- Bifet, A., Castillo, C., Chirita, P.A. and Weber, I. (2005), "An analysis of factors used in search engine ranking", paper presented at the 1st International Workshop on Adversarial Information Retrieval on the Web (AIRWeb), Chiba, available at: <http://airweb.cse.lehigh.edu/2005/bifet.pdf> (accessed April 15, 2018).
- Bridy, A. (2016), "Copyright's digital deputies: DMCA-Plus enforcement by Internet intermediaries", in Rothchild, J.A. (Ed.), *Research Handbook On Electronic Commerce Law*, Edward Elgar Publishing Limited, Cheltenham, pp. 185-210, available at: <https://doi.org/10.4337/9781783479924.00021>
- Brown, B. (2008), "Fortifying the safe harbors: reevaluating the DMCA in a Web 2.0 world", *Berkeley Technology Law Journal*, Vol. 23 No. 1, pp. 437-467, <https://doi.org/10.15779/Z38K68G>

- Calzada, J. and Gil, R. (2017), "What do news aggregators do? Evidence from Google news in Spain and Germany", working paper, Universitat de Barcelona, Barcelona, available at: <http://dx.doi.org/10.2139/ssrn.2837553>
- Cheng, M., Anderson, C.K., Zhu, Z. and Choi, S.C. (2018), "Service online search ads: from a consumer journey view", *Journal of Services Marketing*, Vol. 32 No. 2, pp. 126-141, available at: <https://doi.org/10.1108/JSM-06-2016-0224>
- Chitika Insights (2013), "The value of Google result positioning", available at: <http://info.chitika.com/uploads/4/9/2/1/49215843/chitikainsights-valueofgoogleresultspositioning.pdf> (accessed April 15, 2018).
- Donaldson, C. (2017), "Beyond the DMCA: how Google leverages notice and takedown at scale", *Landslide*, Vol. 10 No. 2, pp. 1-6, available at: www.americanbar.org/content/dam/aba/publications/landslide/2017-nov-dec/beyond-dmca.authcheckdam.pdf (accessed April 15, 2018).
- Dotson, J.P., Fan, R.R., Feit, E.M., Oldham, J.D. and Yeh, Y.H. (2017), "Brand attitudes and search engine queries", *Journal of Interactive Marketing*, Vol. 37 No. 1, pp. 105-116, available at: <https://doi.org/10.1016/j.intmar.2016.10.002>
- Erickson, K. and Kretschmer, M. (2018), "This video is unavailable: analyzing copyright takedown of user-generated content on Youtube", *JIPITEC*, Vol. 9 No. 1, pp. 75-89, available at: <https://doi.org/10.5281/zenodo.1250768>
- Evans, M.P. (2007), "Analysing Google rankings through search engine optimization data", *Internet Research*, Vol. 17 No. 1, pp. 21-37, available at: <https://doi.org/10.1108/10662240710730470>
- Fitzgerald, B., O'Brien, D. and Fitzgerald, A. (2008), "Search engine liability for copyright infringement", in Spink, A. and Zimmer, M. (Eds), *Web Search*, Springer, Berlin and Heidelberg, pp. 103-120, available at: https://doi.org/10.1007/978-3-540-75829-7_7
- Font-Julian, C.I., Ontalba-Ruipérez, J.A. and Orduña-Malea, E. (2018), "Hit count estimate variability for website-specific queries in search engines: the case for rare disease association websites", *Aslib Journal of Information Management*, Vol. 70 No. 2, pp. 192-213, available at: <https://doi.org/10.1108/AJIM-10-2017-0226>
- Fortunato, S., Boguñá, M., Flammini, A. and Menczer, F. (2006), "Approximating PageRank from in-degree", in Aiello, W., Broder, A., Janssen, J. and Milos, E. (Ed.), *Algorithms and Models for the Web-Graph. WAW 2006. Lecture Notes in Computer Science*, Vol. 4936, Springer, Berlin and Heidelberg, pp. 59-71, available at: https://doi.org/10.1007/978-3-540-78808-9_6
- Frantziou, E. (2014), "Further developments in the right to be forgotten: the European court of justice's judgment in case C-131/12, Google Spain, SL, Google Inc v Agencia Espanola de Proteccion de Datos", *Human Rights Law Review*, Vol. 14 No. 4, pp. 761-777, <https://doi.org/10.1093/hrlr/ngu033>
- Granka, L.A., Joachims, T. and Gay, G. (2004), "Eye-tracking analysis of user behavior in WWW search", *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, pp. 478-479, available at: <https://doi.org/10.1145/1008992.1009079>
- Gray, M. (2015), "Copyright Infringement and the first amendment: user-generated content and DMCA interpretation – Youtube. I Tube. We all tube", available at: <http://dx.doi.org/10.2139/ssrn.3163133>.
- Heald, P.J. (2015), "How notice-and-takedown regimes create markets for music on YouTube: an empirical study", in Rahmatian, A. (Ed.), *Concepts of Music and Copyright*, Edward Elgar Publishing Limited, Cheltenham, pp. 195-214, available at: <https://doi.org/10.4337/9781783478194.00013>
- Hotchkiss, G., Alston, S. and Edwards, G. (2005), *Eye Tracking Study: an in Depth Look at Interactions with Google Using Eye Tracking Methodology*, Enquiro Search Solutions, Kelowna.
- Jacques, S., Garstka, K.K., Hviid, M. and Street, J. (2017), "The impact on cultural diversity of automated anti-piracy systems as copyright enforcement mechanisms: an empirical study of YouTube's content ID digital fingerprinting technology", working paper, CREATE, available at: <http://dx.doi.org/10.2139/ssrn.2902714> (accessed April 15, 2018).

- Jansen, B.J. and Resnick, M. (2006), "An examination of searcher's perceptions of nonsponsored and sponsored links during ecommerce web searching", *Journal of the Association for Information Science and Technology*, Vol. 57 No. 14, pp. 1949-1961, available at: <https://doi.org/10.1002/asi.20425>
- Kantar Media (2018), *Online Copyright Infringement Tracker: Latest Wave of Research*, The Intellectual Property Office, Newport.
- Lewandowski, D., Kerkmann, F., Rümmele, S. and Sünkler, S. (2018), "An empirical investigation on search engine ad disclosure", *Journal of the Association for Information Science and Technology*, Vol. 69 No. 3, pp. 420-437, available at: <https://doi.org/10.1002/asi.23963>
- Moreno, L. and Martinez, P. (2013), "Overlapping factors in search engine optimization and web accessibility", *Online Information Review*, Vol. 37 No. 4, pp. 564-580, available at: <https://doi.org/10.1108/OIR-04-2012-0063>
- Mukherjee, P. and Jansen, B.J. (2017), "Conversing and searching: the causal relationship between social media and web search", *Internet Research*, Vol. 27 No. 5, pp. 1209-1226, available at: <https://doi.org/10.1108/IntR-07-2016-0228>
- Pass, G., Chowdhury, A. and Torgeson, C. (2006), "A picture of search", *InfoScale '06 Proceedings of the 1st International Conference on Scalable Information Systems*, Vol. 152, pp. 1-7, available at: <https://doi.org/10.1145/1146847.1146848>
- Peguera, M. (2008), "When the cached link is the weakest link: search engine caches under the digital Millennium copyright", *Journal of the Copyright Society of the USA*, Vol. 56 No. 1, pp. 589-645, available at: <http://noarchive.net/refs/pdfs/cached-link-weakest-ssrn-id1135274.pdf>
- Sag, M. (2017), "Internet safe harbors and the transformation of copyright law", *Notre Dame Law Review*, Vol. 93 No. 1, pp. 499-564, available at: <https://scholarship.law.nd.edu/ndlr/vol93/iss2/2> (accessed April 15, 2018).
- Sen, R. (2005), "Optimal search engine marketing strategy", *International Journal of Electronic Commerce*, Vol. 10 No. 1, pp. 9-25, available at: <https://doi.org/10.1080/10864415.2005.11043964>
- Strzelecki, A. (2017), "IP Address and autonomous system diversification as an important factor for building Google ranking", in Pankowska, M. (Ed.), *Economics and Management in Information Technology Context: International Week*, University of Economics in Katowice, Katowice, pp. 176-185, available at https://web2.ue.katowice.pl/strzelecki/materialy/14strzelecki_VI_int.pdf (accessed April 15, 2018).
- Urban, J.M. and Quilter, L. (2006), "Efficient process or chilling effects-takedown notices under section 512 of the digital Millennium copyright act", *Santa Clara Computer & High Technology Law Journal*, Vol. 22 No. 4, pp. 621-693, available at: <https://digitalcommons.law.scu.edu/chtj/vol22/iss4/1/> (accessed April 15, 2018).
- Urist, J. (2006), "Who'S feeling lucky-skewed incentives, lack of transparency, and manipulation of Google search results under the DMCA", *Brooklyn Journal of Corporate, Financial & Commercial Law*, Vol. 1 No. 1, pp. 209-230, available at: <https://brooklynworks.brooklaw.edu/bjcfcl/vol1/iss1/8/> (accessed April 15, 2018).
- Watters, P. (2013), "Measuring online advertising transparency in Singapore: an investigation of threats to Users", working paper, University of Ballarat, Ballarat, available at: <http://dx.doi.org/10.2139/ssrn.2362626>
- Watters, P.A. (2014), "A systematic approach to measuring advertising transparency online: an Australian case study", *Proceedings of the Second Australasian Web Conference, Australian Computer Society, Inc.*, Vol. 155, pp. 59-67, available at: <https://dx.doi.org/10.2139/ssrn.2362621>
- Watters, P.A., Watters, M. and Ziegler, J. (2014), "Malicious advertising and music piracy: a New Zealand case study", *Cybercrime and Trustworthy Computing Conference (CTC), 2014 Fifth, IEEE*, pp. 22-29, available at: <https://doi.org/10.1109/CTC.2014.13>
- Zhang, P., Stalla-Bourdillon, S. and Gilbert, L. (2017), "A content-linking-context model for automatic assessment of web resources in 'notice-and-take-down' procedures", *The Journal of Web Science*, Vol. 3 No. 1, pp. 1-14, available at: <https://doi.org/10.1145/2908131.2908171>

- Zhang, Y. and Moffat, A. (2006), "Some observations on user search behaviour", *Australian Journal of Intelligent Information Processing Systems*, Vol. 9 No. 2, pp. 1-8, available at: <https://pdfs.semanticscholar.org/27e4/dcd17c6af73477db3c909007fd1a5c2e783e.pdf> (accessed 15 April 2018).
- Zhitomirsky-Geffet, M., Bar-Ilan, J. and Levene, M. (2016), "Testing the stability of 'wisdom of crowds' judgments of search results over time and their similarity with the search engine rankings", *Aslib Journal of Information Management*, Vol. 68 No. 4, pp. 407-427, available at: <https://doi.org/10.1108/AJIM-10-2015-0165>

Corresponding author

Artur Strzelecki can be contacted at: artur.strzelecki@ue.katowice.pl