A Content-Linking-Context Model for Automatic Assessment of Web Resources in "Notice-and-take-down" Procedures

Pei Zhang
School of Electronics and Computer Science
University of Southampton
Southampton, UK
pz2g12@soton.ac.uk

Sophie Stalla-Bourdillon
School of Law
University of Southampton
Southampton, UK
S.Stalla-Bourdillon@soton.ac.uk

Lester Gilbert
School of Electronics and Computer Science
University of Southampton
Southampton, UK
L.H.Gilbert@soton.ac.uk

ABSTRACT

The US Digital Millennium Copyright Act (DMCA) of 1998 [1] adopted a notice-and-take-down procedure to help tackle alleged online infringements through online service providers' actions. The European Directive 2000/31/EC (e-Commerce Directive) [2] introduced similar liability exemptions, but did not specify any take-down procedure. Many intermediary (host, and online search engine) service providers even in Europe have followed this notice-and-take-down procedure to enable copyright owners to issue notices to take down allegedly infringing Web resources. However, the accuracy of take-down is not known, and notice receivers do not reveal clear information about how they check the legitimacy of these requests, about whether and how they check the lawfulness of allegedly infringing content, or what criteria they use for these actions. In this paper, we use Google's Transparency Report as the benchmark to investigate the information content of take-down notices and the accuracy of the resulting take-downs of allegedly infringing Web resources. The analysis of copyright infringement is limited to the five scenarios most frequently encountered in our study of Web resources. Based on our investigation, we propose a Content-Linking-Context (CLC) model of the criteria to be considered by intermediary service providers to achieve more accurate take-down, and investigate technical issues applying the CLC Model to automatically assess web resources and output a 'likelihood of infringement' score.

Keywords

Copyright; Notice-and-take-down; CLC Model; Google Transparency Report.

ISSN 2332-4031; DOI 10.34962/jws-42

1. INTRODUCTION

The emerging Web technologies and online services have brought new challenges to copyright enforcement on the Web [3]. Internet intermediaries such as Internet access providers, content hosts and publishers, and link providers, play an important role in the distribution and communication of online content. They are subject to increasing obligations to monitor allegedly illegal activities undertaken through their platforms, despite the fact there is still a debate over whether, or to what extent, Internet intermediaries ought to have such duties imposed upon them [4]. The DMCA is the first statute to create limitations on the liability of Internet intermediaries on copyright infringement by imposing certain regulatory duties on them. It adopts a notice-and-take-down procedure for host providers and information location tools such as search engines. It requires them to perform several take-down steps when they receive removal notices. In European law, there is no equivalent harmonised procedure being discussed at the Commission level, although similar liability-exemption rules are set forth in the e-Commerce Directive (Articles 12 to 15). Some EU Member States have, however, adopted a notice-and-take-down procedure for copyright infringement [5].

The DMCA does not require intermediary service providers to check the allegedly infringing content to decide whether it is infringing. Instead, it only requires that the content be removed "expeditiously" if the notification substantially complies with Section 512(c)(3). This mechanism has been criticised by many legal researchers because of its major focus on copyright owners' interest and over-protection [6] [7] [8]. Under EU/UK law, it is still unclear whether intermediary service providers have to assess the lawfulness of the allegedly infringing content even in cases in which the allegedly infringing content is not manifestly infringing [9].

In practice, many intermediary service providers such as Google, Twitter and Dailymotion have followed notice-and-take-down procedures. Google has taken a step further to assess take-down requests so as to determine if an infringement has occurred. Because the notice-and-take down procedure implemented by Google for content available in Europe/UK is the same as the one implemented for content available in the US, and because the implementation of the notice-and-take-down procedure by Google has been directly triggered by adopting the DMCA, it makes sense to examine the procedure in the light of the DMCA to fully understand how it works in practice. In order to ensure the accuracy of take-down, it is also important to know the criteria used to examine the allegedly infringing Web resources and the workflow for using such criteria.

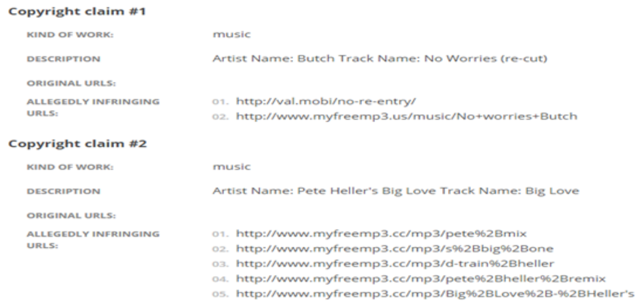
In this paper, firstly, the current take-down practice by online service providers is analysed. Secondly, based on a literature review of legal materials and analysis of current practices, we present a Content-Linking-Context (CLC) Model for copyright related criteria used in assessing content/webpages which are requested to be removed in notices. There are three main components defined in the model. Content is a set of criteria used to compare the similarity between the allegedly infringing work and the original copyright work. Linking is a set of criteria to assess through what method the allegedly infringing work is accessible on a website. Context is a set of criteria to illustrate whether a website is suspected to contain allegedly infringing works. Thirdly, for each criterion in the CLC Model, the background technical implementation to automate each criterion is investigated and an automatic system to dynamically apply the CLC Model to assess Web resources is built. Finally, the CLC Model and the output results of the automatic system are evaluated by experts' review.

2. ANALYSIS OF CURRENT PRACTICES

2.1 State of Claimed Web Resources

To understand more thoroughly the notices and the reported infringing web resources, we analysed the Google Transparency Report, specifically the "request by copyright owners to remove search results"[1], since this report is openly available and provides comprehensive information in respect of webpages associated with potentially infringing content.

According to Google's Transparency Report, 831,185 notices containing over 300 million URLs (used to locate the allegedly infringing content) were received in 2014 in relation to Google Search. Figure 1 shows an example of the copyright claims in a notice. We can see that copyright owners can make several "claims" which contain information about the title, type, and description of the copyright work, original URL, and allegedly infringing URLs.



Copyright claim #1

KIND OF WORK:	music
DESCRIPTION	Artist Name: Butch Track Name: No Worries (re-cut)
ORIGINAL URLS:	
ALLEGEDLY INFRINGING URLS:	01. http://val.mobi/no-re-entry/ 02. http://www.myfreemp3.us/music/No+worries+Butch

Copyright claim #2

KIND OF WORK:	music
DESCRIPTION	Artist Name: Pete Heller's Big Love Track Name: Big Love
ORIGINAL URLS:	
ALLEGEDLY INFRINGING URLS:	01. http://www.myfreemp3.cc/mp3/pete%2Bmix 02. http://www.myfreemp3.cc/mp3/s%2Bbig%2Bone 03. http://www.myfreemp3.cc/mp3/d-train%2Bheller 04. http://www.myfreemp3.cc/mp3/pete%2Bheller%2Bremix 05. http://www.myfreemp3.cc/mp3/Big%2BLove%2B-%2BHeller's

Figure 1. Copyright claims in each notices sent through Web form

We chose one month's notices received by Google dated from September to October 2014. The reason we chose this time period is that our experiment started around the beginning of October 2014, and the latest notice data we could get at that moment was dated from September. For each day, we picked up the first notice received in every hour. And in every notice, two URLs from the first and second claims were selected to make sure the URLs were chosen randomly. In total, 730 URLs were obtained. Among the 730 URLs, 202 pages/content were not found (IP restriction, 404 error[2], copyright work has been removed etc.). The following analysis is based on the 528 pages retrieved.

The URLs point to various types of copyright work. Figure 2 shows the different types of copyright works that were claimed to have been infringed and their percentage in the total of the URLs examined. We can see that Music/Audio represents the largest proportion (57%) of alleged copyright infringing work on the Web. Many websites offer online play functions and supply links for downloading. These music works can be streamed online or downloaded through file sharing websites. At the same time, over half of notices were sent by the right holders in the music industry.

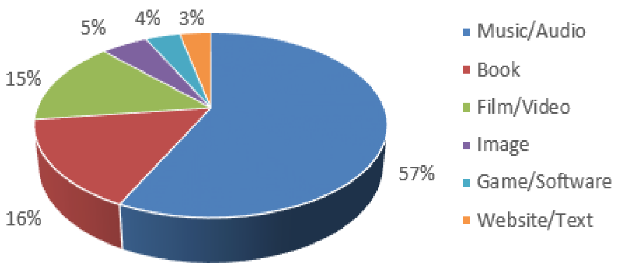


Figure 2. Type of copyright work that claimed to be infringed

There are five types of website which can broadly be said to participate in infringement activities. They are online playing websites, online reading websites, One-Click Hosters [10], index websites, and P2P communities. Figure 3 shows the percentage of different types of reported infringing websites. Online playing websites enable content, including music/audio and film/video, to be played or streamed online. The source could be hosted by the website itself or be embedded from a different host. Most of these websites also offer download function which enable users to download content. The second type of website, online reading websites, applies only to books. Books are displayed in text or image format which allows users to read online freely. The third type is One-Click Hoster sites, such as zippyshare[3], which allows users to upload large files and exchange them by sending corresponding download links to intended recipients of the files. Although this types of websites seems take a small percentage (5%), we believe the number should be much bigger than that. Because there are a large number of One-Click Hoster cases (95) in the 202 URLs which are not calculated into our analysis. The fourth type is websites offering index services. This type of website

[1] https://www.google.com/transparencyreport/removals/copyright/data/?hl=en

[2] http://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html

[3] http://www.zippyshare.com/

searches for content online and indexes corresponding downloadable links. It usually indexes links to different One-Click Hosters. The last type is P2P communities. P2P communities usually supply peer-to-peer download services. The most common P2P services are hosting .torrent files, supplying an index of .torrent files, and running bit torrent tracker servers.

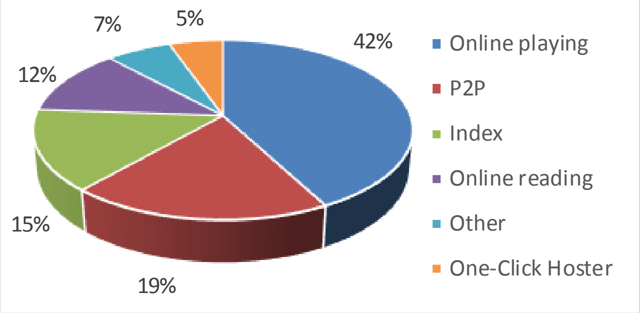


Figure 3. Different types of infringing websites

Figure 4 shows the percentage of different categories of access to copyrighted works. Among the 528 webpages, 249 webpages (47%) are categorized as "Link", which means the infringing sources displayed on the current webpage are hosted by third-party websites from different domains, and the current webpage sets up links for users to view/download copyright work. One hundred and thirty-four webpages (25%) directly host copyright work and have user interfaces which display these works to users, while 82 webpages (16%) are peer-to-peer websites which may host .torrent files, supply indexes of .torrent files, or supply bit torrent tracker servers. We see that, generally speaking, most of the websites analysed do not host copyright work on their own servers, but use a variety of methods to supply links to these works which are hosted on other websites/services.

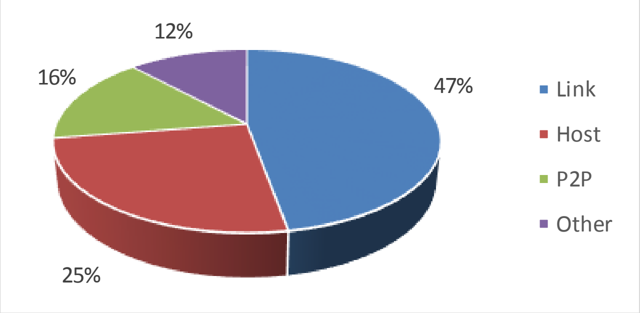


Figure 4. Different ways that copyright work is accessed

2.2 Discussion of Google's Practice on Notice-and-take-down

Google receives a large number of copyright notices every day. Google assesses these notices and the associated URLs to decide whether to remove them. Google releases only simple information about how it assesses take-down requests [11]. One fact known is that Google has adopted a Trusted Copyright Removal Program (TCRP) to help with these assessments. Notice senders who participate in TCRP are believed to be "reliable high accuracy submitters", compared to "non-sophisticated submitters" who issue many "incomplete or abusive" notices [12]. The exact details of the program and how it operates are, however, relatively secret [13]. Seng believes the program is an automated method that allows notice senders to submit large numbers of take-down requests to Google, which Google processes rapidly [14]. No detailed information has been published either about the criteria considered in the decision making process or about how the lawfulness of the content is checked.

A reasonable assumption is that domain-driven analysis plays an important role in the take-down process implemented by Google. From the Google Transparency Report and its website, we can see that Google has been doing extensive data analysis on domain names[4]. The Transparency Report website lists the number of URLs that were reported under the same domain name during a time period, the number of URLs that were already removed under the same domain name, and the number of notice-senders who reported the same domain although they had reported different URLs etc. As a result, the decision to take down is more likely to be according to a top-level domain name suspicion instead of an assessment of the exact content for each URL. Taking the domain vmusice.net[5] as an example, between 8th August 2012 and 8th February 2015, Google received 40,372 notices containing 3,236,150 URLs under this domain. Because vmusice.net is a highly suspicious domain to contain copyright infringing contents, Google's automated program has a high take-down rate of URLs under that domain. The extent to which Google goes further to assess the exact content under each single URL is still unknown. Technically, it is much easier for a system to just compare domains instead of the actual content in the webpages that URLs point to.

From a legal point of view, this method is relatively safe and it follows, to some extent, the practice defined in Section 512(g)(1) DMCA, which indicates that a service provider will not be liable for infringement if the taking down action is based on the "good faith" disabling of access to material that is claimed to be infringing. So if a domain is highly suspected of containing infringing content, intermediary service providers will be acting in "good faith" by removing any URLs under that domain without needing to examine every reported URL. It is arguable whether the domain-driven method is sufficient to ensure reasonable take-down accuracy. Under EU/UK law, there is not a good-Samaritan exemption, which would mean that accuracy is a significant issue.

[4] http://www.google.com/transparencyreport/removals/copyright/domains/?r=all-time

[5] http://www.google.com/transparencyreport/removals/copyright/domains/vmusice.net/, data is captured on 8th February 2015

IPL[6] reviewed Google's take-down procedure and published a report [15] in 2013. It calculated the take-down accuracy for one day (30/03/2013) and predicted the effects an enforced time limit would have on accuracy. The accuracy in that single day was 0.998, and increased to 0.9995 with a longer time limit. However, IPL's definition of accuracy considered a take-down decision to be inaccurate when a removed URL was reinstated. This only happens when Google receives a counter-notice and the number of such notices is very small. Some reasons for this are found in the IPL report and Urban's recent paper [16] where the content provider may be unaware the URL has been taken down, may not understand the law or the counter-notice procedure, or may not be sufficiently interested in the content of the URL to issue a counter-notice.

2.3 Linking Issues on the Web

Linking issues on the Web have triggered a heated debate for legal professionals. An early paper by Deveci [17] believed links bring a number of unresolved issues and raised some copyright concerns associated with linking, such as, "deep linking might bypass advertisements" and "framing might not reveal the ownership of the page called up".

In the US case *Perfect 10, Inc. v. Google Inc* [18], the Ninth Circuit agreed that hyperlinks and framing are not infringing copyright since Google could not "supervise or control" the third-party websites linked to from its search results. It is arguable, however, whether Google would still not be liable assuming Perfect 10 had given Google actual knowledge of specific infringements (e.g. specific URLs for infringing images).

In the *Nils Svensson and Others v Retriever Sverige AB* case [19], an interesting question was raised as to whether hyperlinks are covered by the right to communicate works to the public [20]. The CJEU holds that hyperlinks to protected works which are already freely available online do not infringe copyright. In another case, *BestWater International GmbH v Michael Mebes and Stefan Potsch* [21], the CJEU holds that embedded linking from another freely available website does not constitute an infringement of the right of communication if the work concerned is neither directed at a new public nor communicated by using specific technical means different from that used for the initial communication [22].

In the recent *GS Media BV v Sanoma Media Netherlands BV and others* case [23], the CJEU decided that hyperlinks to a third-party website on which protected works were made available without consent of the rights holder constituted a communication to the public if the person posting those links knew this consent was not given. If the link was posted in the pursuit of "financial gain", the linker was presumed to know about the lack of consent and thus presumed to make a communication to the public.

From a technical perspective, linking is an essential concept on the Web. According to the W3C Recommendation ("HTML5 - A Vocabulary and Associated APIs for HTML and XHTML [7]"), links are a conceptual construct that represents a connection between two resources, one of which is the current document. In HTML, one type of link is linking resources such as CSS or JavaScript files to augment the current web page. Another type of link are hyperlinks which link to other resources that are exposed to the user by the user agent so that the user can use the user agent to navigate to those resources. In this paper we only consider cross domain links and define two types of link depending on where the copyright work is located.

- **Simple link**. A clickable hyperlink which leads users to a new webpage or to a standalone copyright work.
- **Embedded link**. A webpage can embed content from another domain by using a HTML tag. For example, the page "http://example.org/index.html" contains an audio file from http://test.org by using the link <*audio src="http://test.org/music1.mp3"*>. In this example, the audio file from test.org is directly embedded in the webpage "index.html" on example.org by using the HTML5 tag <*audio*> and users are not explicitly notified that the music is from another domain. Content can also be embedded using an <*iframe*> tag. Specifying a URL using the "*src*" attribute in the <*iframe*> tag will direct the browser to fetch the webpage the URL points to and display it in the current webpage. For instance, users can simply add a line similar to this <*iframe width="560" height="315" src= https://www.youtube.com/embed/AbCdEfj frameborder="0" allowfullscreen*> <*/iframe*> to their web page to embed YouTube videos. Similarly, users are not explicitly notified that the content is from another domain. This method of embedding is also called "framing".

3. CONTENT-LINKING-CONTEXT (CLC) MODEL

3.1 Methodology

The objective of the research is to build a Content-Linking-Context Model for accurately analysing copyright infringement on webpages. To build this model we have followed a three-step methodology.

Step one: We undertook a literature review of legal materials from different jurisdictions and current notice-and-take-down practices in order to identify consensual infringement and non-infringement scenarios. Based on this literature review, we constructed five scenarios as listed below: four infringement scenarios and one non-infringement scenario. In order to construct these five scenarios we adopted a conservative view of copyright laws. A conservative view (for a US example see [6]) was needed

[6] http://www.ipl.com

[7] https://www.w3.org/TR/html5/document-metadata.html#the-link-element

to address uncertainties and simplify the analysis. More precisely, we adopted a broad definition of exclusive rights and in particular given the persistence of uncertainties in the field we assumed that even if an act could be considered as being outside the scope of copyright owners' exclusive rights (such as the right to communicate the work to the public), actual knowledge of the presence of infringing material on its system or network on the part of the online service provider (excluding mere conduits) would trigger liability, be it on the ground of copyright liability theories or other liability theories. In addition, we excluded transformative uses of copyright works from our analysis and assumed that partial reproductions of copyright works always amounted to a taking of the originality of the copyright works.

a. Hosting an exact copy of a copyright work without authorization. In this scenario, the website operator hosts the copyright work without the permission of the copyright owner, and usually puts it in the domain of their website for viewing or downloading. We thus assume there is an infringement in this case.
b. Hosting a partial copy of a copyright work without authorization. We define a partial copy of a work as a section of the copyright work which does not have any further additions, and which is a substantial copy. We thus assume there is an infringement in this case.
c. Supplying links (simple or embedded) to an exact copy of a copyright work where making available of the copy is unauthorized. In this scenario, the website operator provides links for users to view/download unlawful content, and the online service provider is informed through notification that the link is to a content, where making available of the content has not been authorised. We thus assume there is an infringement in this case or at the very least, a takedown should happen.
d. Supplying links (simple or embedded) to a partial copy of an unlawful work. This scenario is similar to scenario c, however, instead of giving access to an exact full copy, users are only able to view part of the unauthorized copy. We thus assume there is an infringement in this case or at the very least, a takedown should happen.
e. Supplying links (simple or embedded) to work made publicly available by the copyright owner. We assume there is no infringement.

Step two: In order to investigate whether the most encountered scenarios in practice are covered by the scenarios listed in Step one, we examined the notices in relation to the formats and patterns of reported infringing webpages. From Figure 3, 42% allegedly infringing websites online play content and 12% provide reading content hosted locally or embedded from external websites. All five scenarios created in Step one refer to this type of webpage. Five percent of allegedly infringing websites are host providers, and the number increased to 25% in Figure 4 when online play websites and online reading websites are considered. Scenarios *a* and *b* refer to this type of website, while 47% offer linking services to view/download copyright work (scenarios *c*, *d* and *e)* and 16% provide peer-to-peer content (all five scenarios).

Step three: We derived 3 categories of criteria to be considered in order to determine whether there was an infringement in each of these scenarios and ultimately whether a take-down action would be legitimate. The categorization of content, linking, and context was based on whether the criteria of copyright infringement referred to the website content, the links to it, or the metadata context of the content and the website.

3.2 CLC Model

Our model was limited in the following ways:

1. The model uses the two types of links aforementioned: simple and embedded.
2. The model deals with the five scenarios identified earlier.
3. Only music work is considered in the CLC Model as a starting point, because allegedly infringing music represents the largest proportion of removal requests on the Web (57% in Figure 2).
4. We consider that the principle of exhaustion does not apply to the supply of works online for music. There might be some exceptions in certain systems with regard to certain types of work such as software in the European Union [24], but we assume this is not the case for music. We will therefore not attempt to capture and represent the principle of exhaustion in our CLC model.
5. Although the accuracy of Google's domain-driven method needs further discussion, it does reflect the level of suspicion of a webpage. We use it as a factor to indicate the likelihood of that the webpage contains copyright infringing content.

A Content-Linking-Context Model which contains 12 criteria (C1 to C12) is proposed to indicate different factors we have considered when verifying allegedly infringing web resources in a notice. The model is explained below.

- **Content**. Allegedly infringing content on the webpage to which a URL points needs to be compared with the original copyright work in order to decide on the similarity between them. Criteria C1 and C2 indicate whether the reported content exists on the webpage, and C3 indicates how much the reported content is similar to the original work (by audio comparison).

  C1: URL accessibility. Whether the web resource identified by the URL is still accessible. It is possible that the URL is no longer valid.

  C2: Content existence. When we review the web resource identified by the URL, whether the alleged infringing content can be found on the web page. This criterion co-works with criteria C10 and C11. In the previous 528 webpages analysis, we found 42 webpages contain no music content that was claimed as allegedly infringing content by copyright owners. At the same time, the context information such as the title and performer of the music cannot be found on the 42 webpage either. So from technical perspective, we believe the context information can be used as a first and reliable checking step to identify whether a content is existed or not.

C3: Work (Audio) comparison. If a copy of the work is accessed, its similarity to the original work, whether in whole or part. Both the alleged infringing file and the original copyright music file are used for comparison. There are some technical libraries and open source tools available to compare the two files and give a percentage on how much they match each other.

- **Linking**. Allegedly infringing content could be directly accessed (and played) on the webpage (C4 and C5) or downloadable by users (C6 and C7). Criteria C8 and C9 reflect the requirement that the types of link need to be examined in order to reveal the ownership of the content and whether the source is authorized.

C4: Online access. For music, whether the website offers an online-playing function.

C5: Online playable. Whether the music can be successfully played online.

C6: Download access. Whether the website offers a download function that enables the user to download the music.

C7: Downloadable. Whether the content can be downloaded directly.

C8: Link type of online accessing resources. When an online accessing function is offered, whether the resource is hosted on the current domain, or is embedded from another domain.

C9: Link type of downloadable resources. When a download function is offered, whether the resource is hosted on the current domain, or is linked from another domain for download.

- **Context**. While criteria in Content and Linking can in theory lead to a clear decision of copyright infringement on the Web, in practical instances, however, it may not be so clear. For example, the allegedly infringing music cannot be downloaded or be listened to online when the webpage is viewed (for technical reasons, e.g. temporary broken links), but the decision of taking down by notice receivers still needs to be made. In this case, "Context" information such as whether metadata (C10, C11) of the content appears in the webpage, and whether the host website is highly suspected to contain copyright infringement work (C12), will be used in the decision making process. In addition, if the allegedly infringing content is embedded from/linked to other external website instead of being hosted on the current reported one, C12 assesses whether the external domain is suspected to contain unlawful content.

C10: Title of copyright work. Information about the title of the music.

C11: Performer of the copyright work. Information about the person who performed in the music.

C12: URL suspicion. Google Transparency Report data of URLs that have been claimed to have infringed content is compared to the current URL domain name to find out how many claims have been made under that domain name. This criteria reflects the level of suspicion of a URL.

Figure 5 illustrates the classes and their associations in the CLC model. The Request class represents a removal request and each Request contains one to many WebResources indicated by URLs. The Context consists of criteria about the metadata matching and URL suspicion. The Content class can be either a HostedContent or LinkedContent. LinkedContent means even though the content is displayed within the current WebResource, the content is fetched from a URL other than the URL representing the current WebResource. The TypeOfDelivery class means the content can be delivered by OnlineStreaming, or Downloadable. The LinkedContent will associate with an instance of the Linking class. Depending on the type of the linking, a Linking instance can be one of SimpleLink or EmbeddedLink. Compared with LinkedContent, HostedContent indicates the content delivered is hosted on the current WebResource's URL.

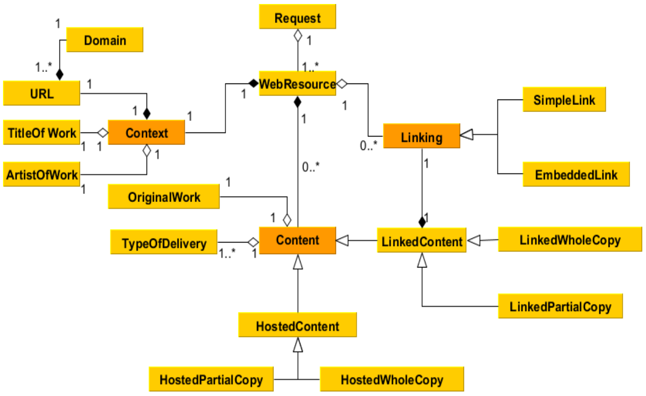


Figure 5. Static Content-Linking-Context conceptual design

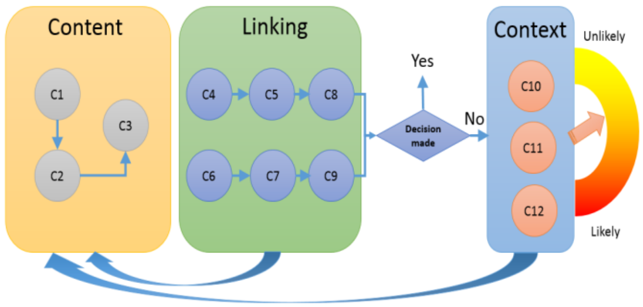


Figure 6. Dynamic Content-Linking-Context illustration

Figure 6 illustrates a dynamic workflow using the CLC Model. If positive answers have been given to C1 and C2 when a removal request is made, allegedly infringing content is compared with original content (C3). At the same time, the Linking criteria identify how the content is displayed (C4, C5, C6, C7) and where the content source is located (C8, C9), so as to further answer the questions of how likely there is a copyright infringement and eventually whether to take it down. In some circumstances, there is no clear answer to copyright infringement by analysing Content and Linking criteria, so C10 to C12 are checked to facilitate any decision on infringement.

4. CLC MODEL AUTOMATION

We investigate the use of the criteria in an automated system to help assess web resources in notice-and-take-down procedures.

4.1 Web page rendering and user interaction

We introduce briefly how a web page (an HTML document) is rendered following a request in a browser and how a user can interact with it (Figure 7).

In this process, except for the HTML markups included in the HTML document, the browser sends requests to the Web server asking for CSS and Javascript files necessary to render the web page.

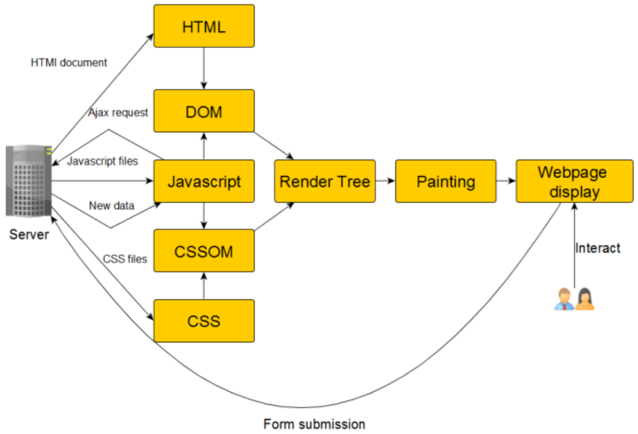


Figure 7. Web page rendering and user interaction

The browser constructs the DOM (Document Object Model) structure according to the HTML markups and the CSS is parsed into the CSS object model (CSSOM). The DOM and CSSOM together determine the Render Tree, layout and look of the web page, which the browser paints as a web page to be displayed to the user. Javascript can programmatically change both the DOM and the CSSOM. This process might be slightly different from browser to browser, but the general steps are the same.

Users view and interact with the displayed web page through the browser. Typical interactions include clicking an element on the web page, using the keyboard to fill a form, etc. Some of the interactions will trigger Javascript on the client side (browser) to modify the web page without contacting the server to handle the interaction, for example, popping up an alert window on the web page, displaying a dropdown menu, changing the font-size, etc. On the other hand there are several kinds of interactions that trigger a request to the server:

1. Form submission: the user fills in a form and submits it.
2. Ajax HTTP call[8]: the user clicks a button or some other page element that sends an asynchronous Javascript HTTP call to the server for new data.
3. Web page redirection: the user clicks a link or some other page element that changes the current URL address and usually results in opening and displaying a completely new page.
4. Multimedia playing: the user interacts with a video or audio player embedded on the Web page and triggers the downloading of multimedia files through browser's native player, or some plugins that can play multimedia resources on the browser (for example, Flash player[9]).

4.2 Criteria automation

Automating the extraction of information relevant to each of the twelve criteria in the CLC Model is discussed in this section.

- **C1: URL accessibility.**

A browser makes an HTTP call to the server for a specified URL. If the URL is not accessible the user sees an error page which may be provided by the server. Technically, the accessibility of a URL can be determined by checking the return code status[10].

Table 1 shows the error codes that reflect C1, according to our analysis of Web resources in section 2.1. Not all 4xx and 5xx HTTP codes are listed in the table because some of them did not appeared in our analysis.

HTTP Code	Code text
400	Bad request
401	Unauthorized
403	Forbidden
404	Not Found
500	Internal server error
502	Bad Gateway
503	Service unavailable
504	Gateway timeout

Table 1. HTTP error codes

There are a number of reasons for an inaccessible URL, and generally we cannot further use the content of the web page containing the inaccessible URL to decide whether there may be content infringement on the requested web page.

To automatically detect the result of C1, we can use regular HTTP client to perform a request to the URL and check the response code. For example, curl[11] is a command line tool in Linux-based operation system to make HTTP call to any

[8] https://en.wikipedia.org/wiki/Ajax_(programming)

[9] https://en.wikipedia.org/wiki/Adobe_Flash_Player

[10] https://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html

[11] https://curl.haxx.se/

server. There are also many UI based tools to help users make HTTP request, such as Postman[12].

- **C2: Content existence.**

Criterion C2 co-works with C10 and C11 and, strictly speaking, the automation of this criterion involves the automation of C4, C5, C6 and C7.

To decide whether content exists, we use text search and comparison functions similar to the "Find" function in the browser. The first step is to extract the text from the HTML document. Then we use an algorithm to match the title or performer text specified in the take down notice with the text extracted from the HTML document.

There are many text extraction tools we can use for the first step, such as Scraper[13], x-ray[14] and IBM Watson's Document Conversion service[15]. For some websites, this is sufficient because the HTML document renders completely including the necessary information. However, some websites use Ajax technology to request additional data from the server after the HTML document has been delivered. In this case, we need to wait until the Ajax calls are finished and the web page is completely rendered before we can analyse its content.

- **C3: Work (Audio) comparison.**

There are many open source and commercial music repositories that we can use to compare the audio file hosted on the allegedly infringing URL to an 'official' repository of music. We use MusicBrainz's open source Fingerprinting service called AcoustID. The service extracts identifying features of music from its original recording and saves these as "fingerprints" into a database. The fingerprint comparison is much quicker than comparing audio files byte by byte.. We download the audio file from the URL and upload it to AcoustID, which gives feedback within a couple of seconds of a match and its percentage similarity.

There are some shortcomings to consider when using the Fingerprinting service.

1. Many recordings are missing in AcoustID, especially in languages other than English, and from less famous performers. The lack of a high similarity match from AcoustID may be because the original recording is missing in the repository.

2. In some take-down notices, the claimed infringement is a sample of the original copyright work. In this case, the fingerprinting is incomplete and the sample cannot adequately be compared with the full original.

- **C4: Online access.**

This criterion is about whether the web page provides access that can lead users to play an audio file online. The technology implementation of actually playing an audio file will be described in C5. If users see a "play" button on the web page, usually it is one of the following implementations:

1. The play button of the HTML native audio and video player. The standard player includes buttons, such as play, volume control, etc., to control the play of the audio or video. There is usually a progress bar to indicate the duration and the remaining time of the audio or video. The play button is usually an icon with a triangle shape arrowhead[16] or similar. The look and feel of the play button can be customised by Javascript and CSS.
2. The play button of video or audio player implemented by a specific technology, such as Flash, Silverlight[17] or other plugin for browsers. Before HTML5, browsers usually needed to install plugins to handle the play of video or audio files. A player can be embedded on the web page and displays player controls.
3. An HTML button or link that will trigger Javascript to play audio or video. This button is different from the play button provided by HTML's native audio or video tag, where the browser handles the default behaviour of the button. Here, the button or link triggers Javascript functions to control the video or audio using HTML5 Media Element API[18]. In some cases, the Javascript can also control play by video/audio plugins.

- **C5: Online playable.**

Following C4 about the web page providing access to an audio file, this section explains the technologies that are commonly used on a web page to actually play video or audio. It must be emphasised that, even though video or audio can be played, it does necessarily mean the web page 'owns' the audio or video file. This is explained in more detail in C8.

A web page that can play audio or video implements one or more of the following techniques.

1. Native HTML <audio> or <video> tags. To specify which file to play, the web page owner specifies the 'src' attribute of the <source> tag within the <audio> or <video> tag.

```
<video width="320" height="240" controls>
  <source src="movie.mp4" type="video/mp4">
  <source src="movie.ogg" type="video/ogg">
```

[12] https://www.getpostman.com

[13] https://scrapy.org/

[14] https://github.com/lapwinglabs/x-ray

[15] https://www.ibm.com/watson/developercloud/document-conversion.html

[16] http://fontawesome.io/icon/play/

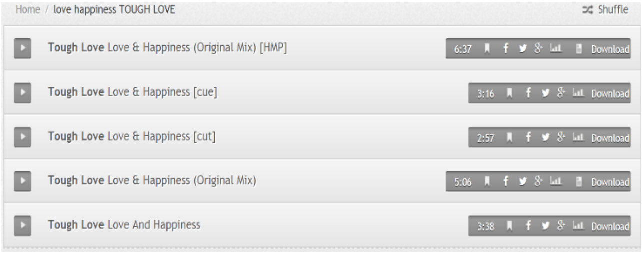
[17] https://www.microsoft.com/silverlight/

[18] https://www.w3.org/TR/html5/embedded-content-0.html#media-elements

</video>

2. Video and audio playing from plugins such as Flash and Silverlight. This type of implementation has been largely discarded in modern web pages because it brings compatibility problem across browsers. However, from our previous analysis, some copy infringement websites still use legacy plugins to play audio. Technically, the plugin is invoked in an HTML document as an <embed> or <object> tag.
3. Video or audio play triggered by Javascript. In this case, there are no <video>, <audio> or <embed> tags, and the Javascript directly provides audio or video play. This may be automatic when opening the web page, or the user clicks a button to trigger the play of a certain file.

The last technique makes the automated detection of C5 very difficult. Unlike the previous two techniques where we can automatically locate the actual file corresponding to a player, the play control of Javascript is totally dependent on how the web page developer programs this function. Figure 8 can explain this situation. Visually, a play button can be related to the playback of the mp3 file in the same row, but an automated process is not able to "see" such visual clues.



Home / love happiness TOUGH LOVE

Tough Love Love & Happiness (Original Mix) [HMP] 6:37

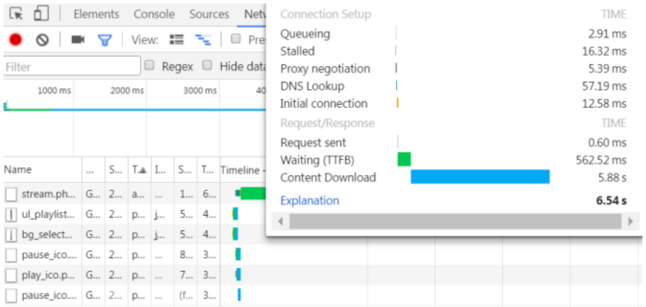
Tough Love Love & Happiness [cue] 3:16

Tough Love Love & Happiness [cut] 2:57

Tough Love Love & Happiness (Original Mix) 5:06

Tough Love Love And Happiness 3:38

Figure 8. My Free MP3 example of javascript controlled audio playback



Name	...	S...	T...	I...	S...	T...	Timeline -
stream.ph...	G...	2...	a...	...	1...	6...	
ul_playlist...	G...	2...	p...	j...	5...	4...	
bg_select...	G...	2...	p...	j...	5...	4...	
pause_ico....	G...	2...	p...	...	8...	3...	
play_ico.p...	G...	2...	p...	...	7...	3...	
pause_ico....	G...	2...	p...	...	(f...	3...	

Connection Setup	TIME
Queueing	2.91 ms
Stalled	16.32 ms
Proxy negotiation	5.39 ms
DNS Lookup	57.19 ms
Initial connection	12.58 ms
Request/Response	TIME
Request sent	0.60 ms
Waiting (TTFB)	562.52 ms
Content Download	5.88 s
Explanation	6.54 s

Figure 9. Network traffic when playback an audio

To determine if audio or video is played after being triggered, we use BrowserProxy to monitor the network traffic. This shows whether an audio or video file is actually sent from a remote server. Using the example of Figure 8, clicking on the first song showns (Figure 9) an audio file with type 'audio/mpeg' being requested from 'stream.php.' Monitoring the network after the play button is clicked shows whether or not the file was actually streamed. Using BrowserProxy to monitor the network traffic is similar to using the networking monitoring functions in Google Chrome's debug mode, where traffic is classified as XHR, JS, CSS, Media, etc. We monitor whether the downloaded packages are classified as Media when a suspect play button is clicked.

- **C6: Download access.**

A recording can be made available on a web page in two ways: (1) the file is played online with an audio or video player, but it is not downloadable or downloaded (i.e. it is streamed); or (2) the file can be downloaded and played offline.

C6 is the starting point for the file downloading criteria group, and it mainly describes whether a download access function is available on the web page, which may lead to the actual download of an audio file. Different from how the file is downloaded (C7 and C9), C6 deals with the website giving users access to a download through one or more steps (or user interactions, such as clicks). A download access function is commonly implemented in one of three ways:

1. An HTML button. Clicking the button submits an HTTP request to the server to initiate file download.
2. An HTML anchor or link. This points to a new URL that initiates file download.
3. Making an HTML element clickable and triggering Javascript to download the file. The Javascript code either relocates the current window to the file or submits a request to the server to download the file.

While the HTML button and HTML anchor are completely different components, modern websites sometimes use CSS and Javascript to make a button seem like a link, and vice visa. For example, Bootstrap, a widely used CSS and javascript framework, defines a 'btn-link' to visually change a button to look like a link. An image on the HTML page can also be an anchor by adding <img> tag within <a> tag. Without inspecting the HTML code directly it can be difficult to decide whether the element corresponding to a download request is a button or a link. However, the code contains text or graphic information indicating that, by clicking the element, a download will be requested, so similar technology to C2 implements the automatic detection of C6.

- **C7: Downloadable.**

If the download access function is available on the web page, a user may follow the download instructions or indications, and this will either lead to an actual download of the file, or failure to download. There might be a possibility that the file cannot be downloaded directly and will become downloadable after a few steps, such as viewing ads, or be redirected to some external website. But the technics to enable the download are roughly the same, which is through specifying HTTP Content-Disposition Header [19] as an

[19] https://tools.ietf.org/html/rfc6266

attachment. The HTTP Content-Disposition Header indicates how the downloaded content should be treated. Possible values are inline, as an attachment, or as a named attachment.

‘Inline’ means the content should be rendered within the current web page, while ‘attachment’ means the file should be downloaded.

We use curl and Postman to detect if a file download happens for a given URL.

- **C8: Link type of online accessing resources**

As discussed in C5, there are many technical ways to play audio and video online. From the visual information on the web page, however, it is usually very difficult to tell which techniques have been used for playback and to tell where the file comes from. A website owner could source the streaming file from a local host or from an external link. As for C6, we monitor the network traffic to decide the type of the link, that is, whether the file is hosted on the current domain, or is embedded from another domain.

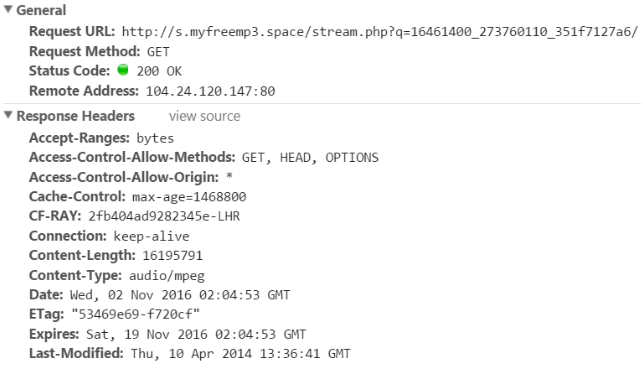
A popular video or audio embedding technology is to use the <iframe> HTML tag, which displays information from another website inside the current website. The URL of the other website is given as the value for the ‘src’ attribute. These two websites can be in different locations and managed by different owners. It is very difficult, and usually impossible, to tell if any component on the website is delivered through an iframe without inspecting the code of the HTML document.

An <iframe> is widely used to provide social features on a website. For example, the Facebook “Like” button on many websites uses an iframe to deliver the button content (image, look and feel, and the id of the liked resource) from Facebook, so the button is not managed by the owner of the website, and it is only a reference. Another example closely related to music copyright infringement is an iframe embedded player from a multimedia sharing website such as YouTube, Vimeo, or SoundCloud. Encountering a player within an iframe makes it quite likely the online access is given by an embedded link.

- **C9: Link type of downloadable resources.**

Following C6 and C7, if an audio file can be downloaded, we examine whether the file is hosted on the current domain, or on another domain. This is done by looking at the HTTP response of the file download, especially the request URL and the remote IP address for the request. Usually, there are two major categories:

1. The file is hosted on the same domain of the claimed URL. In this situation, we can be sure that the website service provider should be responsible for the content of the file.
2. The file is a simple link or streaming address and the content is actually hosted on another domain. In this situation, we can’t be sure whether the file is within the control of the current website provider. Figure 10 shows an example of this case, where the request is the same as that of Figure 9. The MP3 file is streamed from http://s.myfreemp3.space with IP address 104.24.120.147. Even though myfreemp3.space seems similar to the current website, my-free-mp3.com, whether they are managed by the same provider requires further investigation.



```
▼ General
    Request URL: http://s.myfreemp3.space/stream.php?q=16461400_273760110_351f7127a6/
    Request Method: GET
    Status Code: ● 200 OK
    Remote Address: 104.24.120.147:80
▼ Response Headers     view source
    Accept-Ranges: bytes
    Access-Control-Allow-Methods: GET, HEAD, OPTIONS
    Access-Control-Allow-Origin: *
    Cache-Control: max-age=1468800
    CF-RAY: 2fb404ad9282345e-LHR
    Connection: keep-alive
    Content-Length: 16195791
    Content-Type: audio/mpeg
    Date: Wed, 02 Nov 2016 02:04:53 GMT
    ETag: "53469e69-f720cf"
    Expires: Sat, 19 Nov 2016 02:04:53 GMT
    Last-Modified: Thu, 10 Apr 2014 13:36:41 GMT
```

Figure 10. MP3 file hosted from an external domain

Further investigation is also required if content is delivered through the Content Delivery Network, which is a globally distributed server proxy to deliver files faster in different regions, especially for large multimedia files. In this case, the IP address could be masked deliberately by the publisher in order to hide the real IP address.

- **C10 and C11: Matching of the title of copyright work and matching of the performer of copyright work**.

The technology analysis of checking these two criteria is discussed in C2.

- **C12: URL suspicion**.

The Google Transparency Report data of URLs that have been claimed to have infringed content is compared to the current URL domain name to find out the percentage of URLs for that domain which are ultimately removed. The higher the percentage, the higher the URL suspicion value to reflect the likelihood of infringement.

4.3 Automatic assessment system

Figure 11 shows the activity diagram of the automated system to produce a score reflecting the probability of copyright infringement on a web page.

1. The system checks whether the URL is accessible. If the webpage (URL) cannot be accessible as discussed in previous section (negative answers to C1), the system will output the result showing further assessment cannot be completed because of the issue of URL accessibility. This result is indicated as R1 in Figure 11.
2. If the webpage is accessible, C2 (co-works with C10 and C11) is checked to determine whether the content exists on the web page. As explained earlier, the context information is used to inform this determination. A negative answer to C2 terminates the assessment and scores the probability of infringement as 0 (the content

does not exist). A positive answer to C2 leads to a consideration of C6.

3. If neither a download access function nor an online access function can be found on the web page (negative answers to C6 and C4), the assessment terminates and scores the probability of infringement as 0 (the webpage does not supply any method to make the copyright work available or accessible).

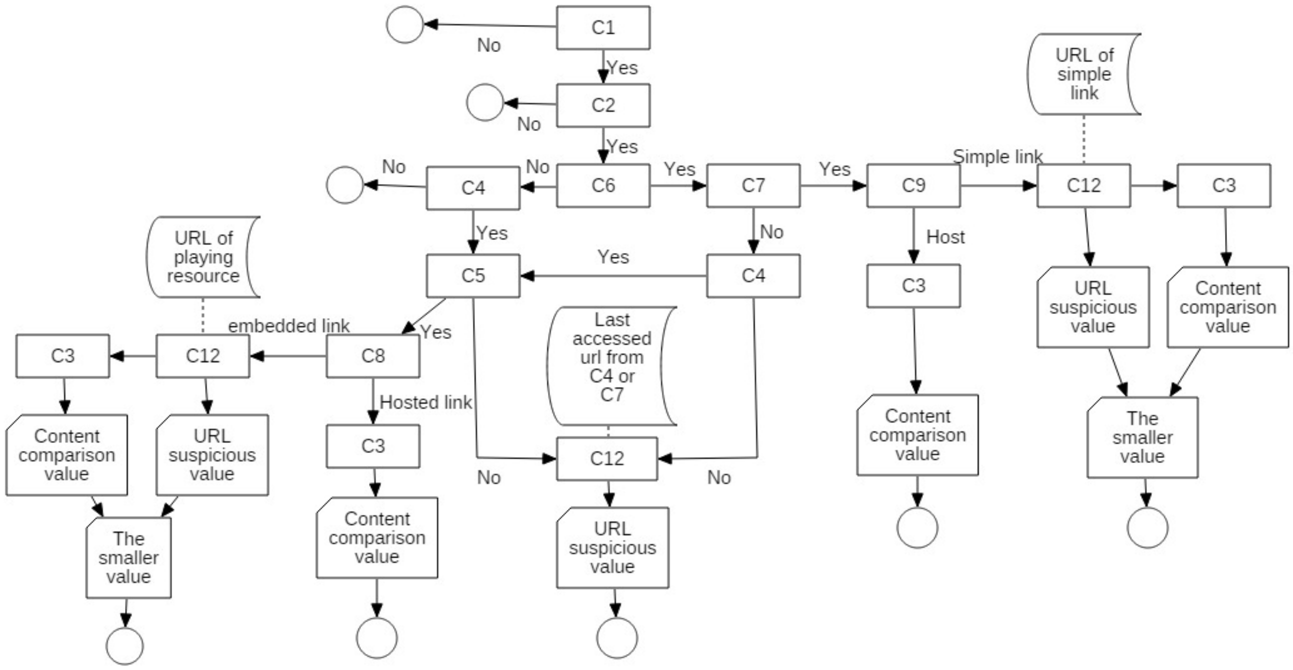


Figure 11. Activity diagram of assessment system

4. If there is no download access function (negative answer to C6), but there is an online access function (positive answer to C4), and the content can be played online (positive answer to C5), whether the content is hosted on the current website or is embedded from external website is checked (C8).
   a) If it is hosted on the current website, the similarity between the content and the original copyright content is checked (C3). The value of the content similarity is the score given for the probability of infringement.
   b) If the content is embedded from an external website, the URL suspicion of the external website (C12) is calculated as well as the similarity between the content and the original copyright content. The score is the smaller of the two values. Where C3 is smaller than C12, although the source where the content comes from is suspicious, the content similarity is lower than this and the probability of infringement is scored accordingly. Conversely, where C3 is bigger than C12, although the content is quite similar to original copyright work, the source is less suspicious and the probability of infringement is scored accordingly.

5. If there is a download access function (positive answer to C6), but the content can neither be downloaded (negative answer to C7) nor be accessible online (negative answer to C4), the download URL's suspicion (C12) is given as the probability of infringement score. Similarly, if only an online access function is found on the webpage (negative answer to C6, and positive answer to C4), but the content cannot be played online (negative answer to C5), online access URL's suspicion (C12) is given as the probability of infringement score.

6. If there is a download access function, and the content can be downloaded (positive answers to C6 and C7), whether the content is hosted on the current website or is linked from external website (C9) is checked.

   a) Similar to step 4a), if it is hosted on the current website, the similarity between the content and the original copyright content is checked (C3). The value of the content similarity is the score given for the probability of infringement.
   b) Similar to step 4b), if the content is linked from an external website, the URL suspicion of the external website (C12) is calculated as well as the similarity between the content and the original copyright content. The score is the smaller of the two values.

5. EVALUATION

In this section, we discuss how the CLC Model and the automatic assessment system are evaluated by expert review.

Thirty URLs were chosen from take-down requests by copyright owners and were given to four lawyers/researchers working in IT/IP law. One webpage was unavailable when the experts came to assess the URLs. The experts completed a questionnaire for each webpage they examined, where they gave their rating on a 5-point Likert scale on how likely the webpage infringed the copyright that was claimed by the copyright owner. This 5-point scale was converted to an infringement score (G1 to G5) as discussed in last section. Following their rating, the experts were shown the criteria defined in the CLC Model and were asked to indicate whether they had used these criteria.

We investigated: 1) whether the criteria defined in the CLC Model were used by experts when these criteria were applicable; 2) whether the pattern of use of each criterion by experts was the same as we expected; and 3) whether experts agreed with the infringement score generated by the automatic system.

While there are 12 criteria defined in the CLC Model, the evaluation questionnaire presented 9. All the webpages given to the experts existed, so C1 was not presented. Because C2 co-works with C10 and C11, and both C10 and C11 are always used as a first step to check content existence, C2 was not presented. C8 and C9 relate to the technology of the link type (host, simple link, embedded link) with which the experts may not be familiar, but they may use a criterion about the source of a copyright work. Criteria C8 and C9 were combined into one in the questionnaire.

Table 2 shows the pattern of criteria usage. For each criterion, the total number of uses by all the experts when they assessed 29 URLs was recorded. When a criterion was applicable to assess a particular URL, the criterion was expected to be used by experts, and the probability of use was calculated.

Criterion	C3	C4	C5	C6	C7	C8-C9	C10	C11	C12
No. of uses by experts	52	74	62	76	48	56	86	85	85
Expected No. of uses	60	77	65	77	56	60	86	86	86
Prob-ability of use	.87	.96	.95	.99	.86	.93	1.0	.99	.99

Table 2. Pattern of criteria usage

The probability of use of each criterion, when applicable, was between 0.86 and 1, indicating the criteria defined in the CLC Model were frequently used by experts. Whether the pattern of use of each criterion by experts was the same as we expected was assessed by a Chi-squared test. The result was $\chi^2= 2.77, df = 8, p >> 0.05$, suggesting that the pattern of use of the criteria by experts was the same as we expected.

To investigate whether the scores generated by the automatic assessment system agreed with the expert ratings, the correlation [25] between the system's score and the experts' rating was calculated. Table 3 shows the numbers. The infringement score generated by the automatic system was significantly correlated with the experts' rating (r = 0.537, df = 27, p= 0.003), suggesting that when the experts give higher ratings on infringement, our system similarly gives higher scores.

URL No.	System score	Experts rating	URL No.	System score	Experts rating
1	.976	4.7	16	.000	3.7
2	.000	4.7	17	.945	4.3
3	.994	4.0	18	.964	4.3
4	.912	5.0	19	.000	4.0
5	.000	3.7	20	.001	3.3
6	.875	4.0	21	.000	2.3
7	.001	2.7	22	.444	2.7
8	.972	4.3	23	.986	4.3
9	.000	5.0	24	.842	5.0
10	.930	4.7	25	.946	4.7
11	.017	4.7	26	.444	3.3
12	.000	1.7	27	.000	2.3
13	.915	5.0	28	.444	3.0
14	.017	1.3	29	.120	4.3
15	.966	4.3			

Table 3. System score and experts rating for each URL

6. CONCLUSION AND FUTURE WORK

How to reform the notice-and-take-down procedure is hotly debated by legal professionals. Applying proper criteria to assess Web resources in removal requests in order to support notice receivers' decision making process is essential to improve the procedure. We designed a CLC Model to represent 12 criteria and indicate how these criteria operate for the analysis of allegedly infringing Web resources.

The purpose of CLC Model is to help verify copyright infringing activity on webpages, preferably in an automatic manner. In consequence, we developed a system to apply the CLC Model and automatically assess web resources and generate analytic results. Strictly speaking, only Judges are properly placed to make a decision on the lawfulness of a Web resource, so the output of the system is a score to indicate the likelihood of infringement with a view to support the decision making process and not replace it.

In the CLC Model, it is difficult to fully automate some criteria. Given the variety and fast development of Web technologies used to present a Web page, we need more automatic and evolving mechanisms to detect the content and different components on Web pages. So as a supplement to using the Web page information extraction and monitoring technologies proposed in Section 4, computer vision and machine learning technologies can be used in the future work to recognise the existence of some Web components in CLC model. A machine learning model will be developed which will take screenshots of different instances of a Web page as input, and analyse whether components such as a video player, play buttons, download buttons, login forms, etc., are likely to be presented to viewers.

While the correlation between experts' rating and the system's score was significant and reasonably substantial, gathering labels of more URLs from experts and developing machine learning algorithms to categorise the Web pages will be the next step to further validate the model and the technology implementation. However, obtaining labelled data from experts and using machine learning methods are not the main concerns in this paper. Firstly, there are major difficulties in gathering more labelled data because most of the suspected infringement Web pages are not stable, and can and are taken offline quite quickly. Secondly, as mentioned in Section 2.1, hundreds of thousands of URLs are requested to be viewed and examined for takedown every day by Google, for example. Even if machine learning techniques are going to be used, the real technical challenge is how to efficiently extract feature values from such large numbers of Web pages. This paper proposes crucial steps and methods for automatic feature value extraction which will provide valid data for later use as training data in machine learning.

The CLC Model and the automatic system could be used by online link providers, such as search engine providers and index service providers. In the future they could also be adopted by anti-piracy service providers such as Muso, Degban, and AudioLock.Net in order to help them filter allegedly infringing websites when they send out take-down notices.

7. REFERENCES

[1] *Digital Millennium Copyright Act, H.R. 2281, 105th Congress*. 1998.

[2] *Directive 2000/31/EC of the European Parliament and of the Council on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (June 8, 2000), OJ L 178, 17.7.2000, p. 1–16.* .

[3] N. Elkin-Koren, "After Twenty Years: Copyright Liability of Online Intermediaries," *Evol. Equilib. Copyr. Digit. Age (Susy Frankel Daniel J Gervais eds.)(2014 Forthcoming)*, 2014.

[4] S. Stalla-Bourdillon, "Online monitoring, filtering, blocking... what is the difference? Where to draw the line?," in *International Association of IT Lawyers*, Copenhagen,DK: International Association of IT Lawyers, 2012.

[5] "First Report on the application of Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market," 2003.

[6] J. M. Urban and L. Quilter, "Efficient Process or Chilling Effects-Takedown Notices under Section 512 of the Digital Millennium Copyright Act," *St. Cl. Comput. High Tech. LJ*, vol. 22, p. 621, 2005.

[7] J. H. Reichman, G. B. Dinwoodie, and P. Samuelson, "Reverse Notice and Takedown Regime to Enable Pubic Interest Uses of Technically Protected Copyrighted Works, A," *Berkeley Tech. LJ*, vol. 22, p. 981, 2007.

[8] J. Cobia, "Digital Millennium Copyright Act Takedown Notice Procedure: Misuses, Abuses, and Shortcomings of the Process, The," *Minn. JL Sci. Tech.*, vol. 10, p. 387, 2008.

[9] S. Stalla-Bourdillon, "Sometimes one is not enough! Securing freedom of expression, encouraging private regulation, or subsidizing Internet intermediaries or all three at the same time: the dilemma of Internet intermediaries' liability," *J. Int. Commer. Law Technol.*, vol. 7, no. 2, 2012.

[10] T. Lauinger, M. Szydlowski, K. Onarlioglu, G. Wondracek, E. Kirda, and C. Kruegel, "Clickonomics: Determining the Effect of Anti-Piracy Measures for One-Click Hosting.," in *NDSS*, 2013.

[11] Google, "How Google Fights Piracy," 2013.

[12] R. Tushnet, "PTO/NTIA: notice and takedown-Improving the Operation of the Notice and Takedown System," 2013. [Online]. Available: http://tushnet.blogspot.co.uk/2013/12/ptontia-notice-and-takedown.html.

[13] M. Leiser, "The copyright issue and censorship threat buried within Google's transparency report," 2013. [Online]. Available: http://www.thedrum.com/news/2013/12/23/copyright-issue-and-censorship-threat-buried-within-googles-transparency-report.

[14] D. Seng, "The State of the Discordant Union: An Empirical Analysis of DMCA Takedown Notices," *Virginia J. Law Technol. Forthcom.*, 2014.

[15] IPL, "A report by IPL for Google - Modelling the takedown process," 2013.

[16] J. M. Urban, J. Karaganis, and B. L. Schofield, "Notice and Takedown in Everyday Practice," *Available SSRN 2755628*, 2016.

[17] H. A. Deveci, "Hyperlinks Oscillating at the Crossroads," *CTLR-OXFORD-*, vol. 10, no. 4, pp. 82–94, 2004.

[18] *Perfect 10, Inc. v. Amazon.com, Inc. and A9.com Inc. and Google Inc. 508 F.3d 1146 (9th Cir. 2007).* 2006.

[19] *CJEU C-466/12 Nils Svensson et al v Retriever Sverige AB, 13 February 2014 ECLI:EU:C:2014:76..*

[20] E. Arezzo, "Hyperlinks and Making Available Right in the European Union--What Future for the Internet After Svensson?," *IIC-International Rev. Intellect. Prop. Compet. Law*, vol. 45, no. 5, pp. 524–555, 2014.

[21] *CJEU C-348/13 BestWater International GmbH v Michael Mebes and Stefan Potsch of 21 October 2014 ECLI:EU:C:2014:2315..*

[22] E. Rosati and O. Löffel, "That BestWater order: it's up to the rightholders to monitor online use of their works," 2014. [Online]. Available: http://ipkitten.blogspot.fr/2014/10/that-bestwater-order-its-up-to.html. [Accessed: 12-Dec-2015].

[23] *CJEU C-160/15 GS Media BV v Sanoma Media Netherlands BV, Playboy Enterprises International Inc., Britt Geertruida Dekker, 8 September 2016 ECLI:EU:C:2016:644..*

[24] *CJEU C-128/11 Usedsoft GmbH v Oracle International Corp, 3 July 2012 ECLI:EU:C:2012:407..*

[25] A. Field, *Discovering statistics using IBM SPSS statistics*, 4th ed. Sage, 2013.