

IP: Google's copyright breaches

A decision in a US court could have widespread repercussions for the search engine industry as Perfect 10 shows imperfections in how data is used by third parties.

There are search engines and there are electronic libraries and services such as google.com have blurred the distinctions. Then there are services such as archive.org which provide an undoubted public service but tread on dangerous ground when it comes to the rights of those who place original material on the web.

For copyright law, in principle, protects the originator of artistic works, which include both text and images, from the simple ripping off of their material.

Google not only indexes material, but it allows users to access its "cache" - the stored copy of webpages. It also has a service called Image Search which downloads images from the websites visited by its "googlebots" and displays reduced size "thumbnail" versions of the images.

Perfect 10, a US girlie magazine, told Google that it was not allowed to place thumbnails of its images in the Image Search service. Google claimed that it has a right to so do under the so called "fair use" exemption to US copyright law.

The Court held that Google was not permitted to download and cache or place in the Image Search service copies of Perfect 10's images. However, the Court denied Perfect 10's claim that Google should not be permitted to index Perfect 10's images where the images were hosted - without permission - on a third party's site.

The case isn't over: yesterday's decision was a step towards an anticipated injunction to be awarded to Perfect 10 against Google, but at an early stage in the proceedings.

The decision opens a can of worms, for if storing and displaying an image is a breach of copyright, then so must be storing and displaying whole web pages, which is what the cache service run by Google, Yahoo and others does. The argument will be over what amounts to fair use. There is no doubt that fair use would apply to the actual search results. But that falls a long

way short of making the whole page available.

Archive.Org stores historical copies of millions of websites, going back over several years. It's a very valuable tool for looking at the history not just of websites as such but equally importantly what companies, for example, have said about themselves over time. For investigators, Archive.Org often presents information about a business that the business no longer makes public.

It is in this area that caching raises a complex issue: if a publisher wishes to "unpublish" something, then simply removing it from his own website won't be enough to remove all copies from the internet. This is, of course, especially important in relation to items that have caused some sort of offence. The search companies are, in this way, putting themselves in a dangerous position for they become, in effect, republishers. If a publisher takes down material but the search engines continue to display it, then their principle defence to such a claim would be that their business processes do not update pages quickly enough.

A further problem is in relation to webpages that are updated: a list of people missing in a disaster may be cached on Monday, updated on Tuesday but the website becomes overloaded. If relatives read the cache, they may be affected by good or bad news, that is no longer current. If they forget that they accessed the cache rather than the actual site, as one suspects they probably will, they are likely to criticise the operator of the site.

Web analysis company Alexa.Com, owned by Amazon, places reduced size images of the front page of websites alongside the information it displays about sites.

Google is saying that, if an injunction is granted, it will apply only to Perfect 10's images. Strictly that is true. But it's a childish response in the land of the class action. It may be only a matter of hours before someone launches a class action and invites millions of website publishers to try to get a piece of the Google pie, and others too.

Of course, Google will say that control rests with the site owner who can insert a line into the first page to prevent robots accessing the site. But that denies the reality that it presumes that search-bots have the right to access sites unless specifically told not to. That's a debateable argument as, increasingly, webspace is seen as akin to physical space. We are in the realms of trespassers and invitees when that discussion gets under way. And there is another much more fundamental issue: failure to appear in google, yahoo, etc. is the kiss of death to a website. So blocking the bots is commercially undesirable.

The real question is much more simple: must websites that do not block search-bots have to accept the whole package, or can they select what data and images they allow the search engine to display?

This is, ultimately, at the heart of the issue and it's a very important question. If we, for example, allow a search-bot access to our members areas so as to index content, we do not want that content cached and therefore available free to the world. In that case, however, there is a distinction as against the previous concept for in this case we would be specifically granting access, and therefore would probably be deemed to be accepting the risks associated with it.

The case may seem, right now, to be about who has the right to show naked women. But it is one of the most important cases about the internet yet to come to court.