A few legal comments on *spamdexing*

by Gerrit Vandendriessche[1]
Altius, Brussels

What is *spamdexing?*

The Internet contains a lot of information. In 2002, the total number of web pages was estimated at 2.024 million; by 2005 this estimate had risen to 11.5 billion[2]. The ever-increasing amount of information found on the Internet also has disadvantages. In this jungle of information, it is getting harder and harder to see the wood for the trees. Internet search engines have become vitally important aids to find information in a fast and efficient way. Without them, it is very hard to find information on the Internet, unless the user has memorised the *Unique Resource Locator (URL)* of the website in question. For content providers, it is also getting harder, in the face of competition from other websites, to communicate their messages to the general public. Since Internet search engines are the *de facto* gateways to information on the Internet, websites battle to get the highest possible rankings in the search results. Setting high rankings is essential since research has shown that surfers tend to only look at (and click on) the first ten search results[3]. When a content provider wants his or her website content to be seen by as many people as possible, he or she will have to ensure that the site is ranked as high as possible in the search results and preferably within the first ten search results.

Previously, content providers achieved this by manipulating the *meta tags* in the source code of their websites. However, the search engines successfully countered this manipulation and nowadays the bigger search engines no longer index websites using meta tags alone[4]. Their search robots' search algorithms have been developed and made more complex, which means that, to manipulate search results today, content providers have to use *search engine marketing or search engine optimisation* ('*SEO'*).

SEO can be defined as giving better visibility to a website via Internet search engines. *SEO* mainly focuses on two areas: natural search results and commercial search results[5]. Natural search results are those generated by search robots using non-commercial indexation criteria. Their content and the order in which they are displayed cannot be

[1] The author is lawyer at the Brussels bar and works for the Altius legal practice (www.altius.com).

[2] Source: Wikipedia, referring to http://www.netz-tipp.de/languages.html and http://www.cs.uiowa.edu/~asignori/web-size/.

[3] This is scientifically studied by W.Wirth, T. Böcking, V. Karnowski and T. van Pape in *"Heuristic and Systematic Use of Search Engines",in Journal of Mediated Communication*, 12, 2007, p778.

[4] Google apparently no longer takes *meta tags* into account when indexing websites, but search engines like MSN and Yahoo continue to do so, to a limited extent.

[5] On some websites the term *SEO* is used to describe the manipulation of natural search results and *search engine marketing* for the manipulation of commercial search results.

influenced by payment (by advertisers or other parties)[6], in contrast to commercial (or sponsored) search results, whose content and order are determined commercially, most often by payment. When using the bigger search engines, users will notice that the natural and commercial search results appear separately and the commercial search results are clearly indicated as such. In this article, I will not go further into the manipulation of commercial search results, though there is an increasing amount of case law on this subject[7].

Marketeers prefer *SEO* of natural search results. Experienced users are more interested in natural search results than in commercial search results. To be able to manipulate natural search results, one has to understand how search engines index web pages. The search algorithms of the larger search engines are top secret, and are each different.

Spamdexing[8] is a form of *SEO* and a generic term for techniques that aim to improve the position of a website in natural search results by attempting to manipulate the indexing criteria used by search robots. Spamdexing techniques include *hidden text*, *keyword stuffing*, *doorway pages*, *link farms*, *hidden links*, *mirror websites*, *cloaking,* etc. (see below).

The Internet search engine providers do not like spamdexing because it “falsifies” natural search results. That is the reason why most of them forbid spamdexing in their guidelines. When some search engines detect spamdexing, the websites concerned are *blacklisted* and are not displayed in search results until the spamdexing has stopped.

***Spamdexing* in the relationship between competitors**

First, let us examine the situation where web page A is the victim of spamdexing by web page B, belonging to its competitor. The order of the natural search results is changed so that web page B suddenly appears higher in the search results, often above web page A. The question raised here is what legal action the owner of web page A can take against the owner of web page B or the spamdexing service provider.

Of course, changes in the order of search results are not necessarily the result of spamdexing: there may be other causes, including changes to the search engine’s algorithm, or changes to the web page. Changes in search results only become litigious when they are caused by an unlawful act.

[6] The degree to which natural search results are really “natural”, cannot be checked because the search engines’ search algorithms are secret. Some smaller search engines have been known to present commercial search results as “natural”.
[7] Most of the disputes in this area have been related to Google *Adwords.*
[8] *Spamdexing* is a contraction of *spamming* and *indexing.*

This unlawful act could be either of the two commonest forms of spamdexing: (i) manipulation of the source code of a web page leading it to be ranked higher in the search results, or (ii) using other web pages to achieve a higher ranking for the web page.

As the search engine providers become more aware of spamdexing and modify their search criteria and algorithms, it is likely that new forms of spamdexing will emerge in the future.

Spamdexing of the first type is a variation on the manipulation of *meta tags*. *Meta tag* manipulation adds certain words, mostly trade names or trade marks used by competitors, to the *meta tag "keyword"*[9] in the web page's source code so as to improve its position in the search results, mostly to the disadvantage of competing websites. Most search engines can now detect *meta tag "keyword"* manipulation. *Spamdexing* therefore relies on other, more sophisticated techniques, including *hidden text* and *keyword stuffing*.

Hidden text is the inclusion of words or sentences on a web page in the same colour as the background and using a very small (unreadable) font[10]. While it is hidden and invisible to the visitors to a webpage, the text is visible to search engine robots via the web page's source code, and will thus affect the ranking of the web page in the search results.

In *keyword stuffing,* the same keyword is repeated several times in different places on the web page. Smaller search engine robots rank pages according to the number of times a keyword appears, so this can result in an improved ranking position. However, larger search engine robots are now able to detect the excessive use of keywords on a web page. Several techniques of this first type of spamdexing can be combined.

In the second type of spamdexing, website B has one or more accessory web pages. This type of spamdexing is based on the degree to which search engines take the popularity of a web page into account when calculating search result rankings. If a large number of web pages contain hyperlinks to one particular web page, this page will achieve a higher ranking in search results[11]. Spamdexing exploits this by simulating the popularity of a web page, by creating new web pages whose only function is to hyperlink to the target web page. For example, *a link farm* is a collection of web pages that link to each other for the sole reason of creating mutual popularity. Sometimes these artificial hyperlinks are also invisible (*hidden links*). *Sybil attack* is the technique by which one person, using different simulated identities, creates a number of websites that link to each other. Web pages with little or no content, whose primary function is to link to another web page to improve search result rankings are referred to as *gateway pages* or *doorway pages.* Sometimes users are automatically redirected to another web page than the one they click on in the search results (this is

[9] Other kinds of *meta tags* also exist, such as the *meta tag "description*".

[10] There are other techniques to include words in the html code of a web page in such a way that they are not visible to users, but, due to their highly technical nature I will not discuss them in this article.

[11] This technology was used first by Google, under the name of PageRank, named after Mr Page, one of Google's founders.

called *URL redirection*). Another technique, called *cloaking,* displays to search engines a different version of the web page than the one displayed to users. One form of *cloaking* is *code swapping*, where, once a web page has obtained a high ranking in search results, it is replaced by another web page.

As far as the first type of spamdexing is concerned*,* from a legal point of view, the first question concerns the “invisible” text. If such text uses the registered trade mark or brand name of a third party without permission, then a legal claim based on trade mark law, possibly combined with unfair trade practice rules (for example misleading or confusing advertising, or unlawful comparative advertising) may be possible. Hidden text can be challenged by reference to existing case law on the use of trade marks in *meta tags*. If the “invisible” text is a (partial) copy of another web page, then copyright becomes a consideration.

Both forms of spamdexing could also be seen as misleading advertising, regardless of any infringement of intellectual property law. A web page on which a company presents, and *a fortiori* offers, its services or products could be considered as an “information society service”. It can be argued that some of the spamdexing techniques described above constitute advertising for an “information society service” (i.e. a website) that intends to mislead users as to the characteristics of a web page by using *hidden text, doorway pages, cloaking, or URL redirection,* or as to the web page’s popularity, by using *link farms, hidden links, or sybil attacks*. In short, there is deception of users by altering the search engine’s natural search results. The artificially upgraded position in the search results imputes characteristics to a website that makes users believe that the web page will answer their query.

Spamdexing can also be challenged based on unfair trade practice rules. Under these rules, spamdexing would qualify as a practice prejudicing the professional interests of other vendors. There is no doubt that spamdexing prejudices the professional interests of other website owners, as the lower position of their web pages in search results directly results in fewer visitors, meaning fewer customers and possibly also reduced advertising revenues from banner clicks. In order to successfully sue under trade practice rules, the claimant would have to prove that spamdexing conflicts with the sector’s normal standards of conduct. Most search engines have issued rules and guidelines that forbid spamdexing. Certain *SEO* companies have also denounced spamdexing. These rules and guidelines could be seen as standards of conduct, although some reservations must be made. Most search engines forbid current spamdexing techniques, but fail to define them precisely. Some techniques (for example *cloaking* and, more particularly, *code swapping*, and *hidden links*) are obviously intentionally abusive and deceptive and are likely to constitute an unfair trade practice. Other techniques (e.g. *hidden text* and *URL redirection)* could, under certain circumstances, be justified for technical reasons. It is also difficult to distinguish “legitimate” from “illegitimate” hidden text. Other techniques need to be interpreted in light of their actual effects. For example, when examining *keyword stuffing,* how many repeats of a certain keyword are required for it to be deemed an unfair trade practice? How many mutuallylinked pages are needed to constitute a *link farm* or a *Sybil attack*? Case law will undoubtedly have to elaborate on this.

The EU Directive on unfair trade practices[12] could provide another ground for combating spamdexing, as, “deception” no longer only applies to advertising, but also to business practices. Spamdexing techniques using invisible means are no longer immune from sanction, because these can be qualified as a commercial practice which “*contains false information and is therefore untruthful or in any way, including overall presentation, deceives or is likely to deceive the average consumer, even if the information is factually correct”.*

Spamdexing in the relationship between website owners and search engines

Spamdexing not only harms competing websites, but also the professional reputations of search engines because it degrades the value of the natural search results: users do not really find what they are looking for and have the impression that the natural search results are not all that “natural”. They are therefore less likely to use the search engine in the future, and a drop in user volume affects the search engine’s advertising revenues. The search engine also loses revenue because spamdexing allows website owners to achieve a high position in search results , which will make them less inclined to pay for commercial search results.

The most immediate practical sanction available to search engines to counter spamdexing is blacklisting (the removal of the web page from the search index) until the spamdexing is removed, followed by legal action against recidivist or large-scale spamdexing as required, based on one or more of the above-mentioned legal grounds.

Proof of spamdexing

To take legal action against spamdexing, a claimant has to take great care to furnish sufficient proof. First of all, the key words and the resulting search results have to be officially recorded (by a bailiff or similar person). Certified print-outs of the relevant web pages (both the displayed and the source-code versions) will have to be obtained. For the second type of spamdexing mentioned above, both versions of the “accessory” web pages will also have to be officially recorded[13]. To detect and prove *cloaking,* a comparison must be made between the web page displayed when clicking on the hyperlink to it in the search results and the web page displayed when clicking on the hyperlink to the cached version provided by certain search engines. Should comparisons over a longer period be required, historical versions of web pages can be found in the Internet Archive[14].

[12] *Pub. L.*, 149/22, June 11, 2006

[13] To find out which web pages link to one specific web page, give the command “link to: [name of the website]” in the search engine. All the web pages that contain a hyperlink to the web page concerned will then be displayed.

[14] See the “*Waybackmachine*” on www.archive.org.

Conclusion

Spamdexing has been around for some time, and, as search engine use increases, spamdexing techniques are likely to be used more often in the future. The victims of spamdexing can invoke rules prohibiting misleading advertising and/or unfair trade practices.

Spamdexing techniques are usually used by website developers or *SEO* companies, paid by website owners to improve the search results of their websites. Although these companies are mostly legitimately adding value, I would advise website owners to arm themselves contractually against over-assiduous *SEO* companies that exceed what is generally accepted in the sector. The consequences of blacklisting by search engines can be far-reaching: from one day to the next a web page disappears from the search results of one or more search engines. It is therefore recommended to stipulate in a contract with *SEO* companies that spamdexing is not to be used, that the SEO company has to abide by the rules of the search engine owners and that the party hiring the SEO company cannot be held liable for claims for damages resulting from spamdexing.

In conclusion, I would stress that not all *SEO* activities and techniques are illegitimate. We must take care not to throw out the baby with the bathwater. *SEO* is not a *de facto* illegitimate activity, indeed most search engines offer hints on how to improve web page indexation. It is only when an *SEO* company resorts to spamdexing techniques in a deceptive and abusive way, without any technical reason, with the intention of manipulating search results, that its behaviour can be considered illegitimate.

Gerrit Vandendriessche
Altius, Brussels
October 24, 2007
gerrit.vandendriessche@altius.com