

DES « ROBINS DES BOIS » LIBÈRENT LES LIVRES DE GOOGLE BOOKS

LE 13 OCTOBRE 2010 LIONEL MAUREL (CALIMAQ)

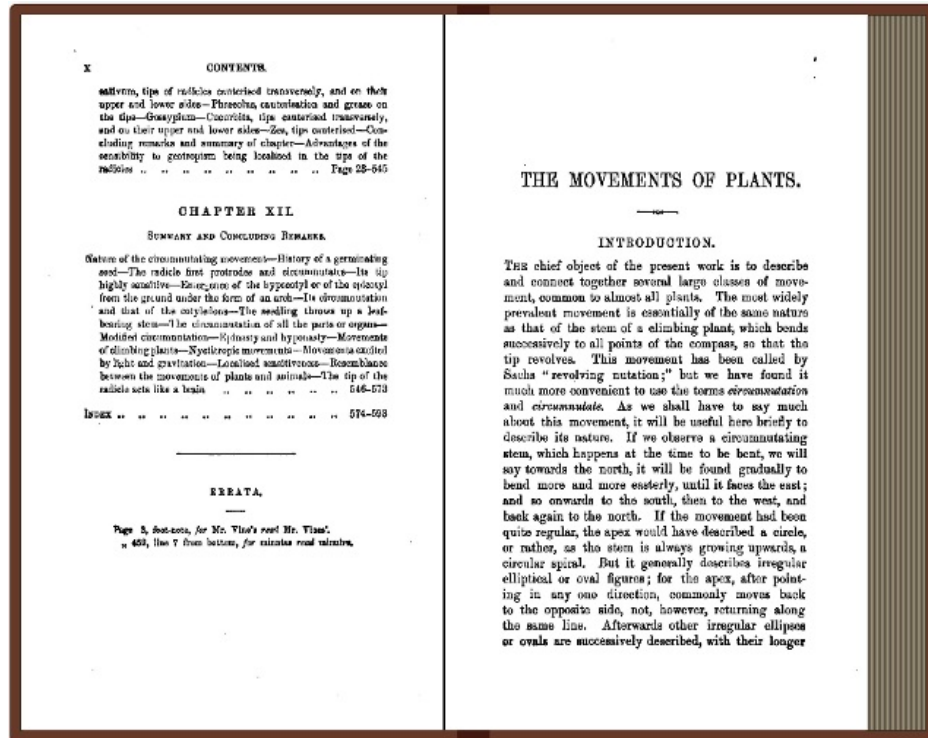
Des internautes ont téléchargé plus de 900 000 livres sur Google Books pour les mettre sur Internet Archive, supprimer les restrictions d'usage imposées par Google et les remettre pleinement dans le domaine public.

Hier, sous le **billet que j'avais consacré cette semaine aux livres du domaine public dans Google Books**, un internaute nommé Pablo Iriarte a laissé un **très intéressant commentaire**, qui m'incite à revenir sur cette question dans un nouveau billet.

Visiblement, les éditeurs ne sont pas les seuls à essayer de récupérer des ouvrages scannés par Google pour les réutiliser à leurs propres fins. Des utilisateurs d'**Internet Archive** auraient procédé à des téléchargements massifs (900 000 livres !) pour les charger dans Internet Archive, afin qu'ils puissent être débarrassés des restrictions d'usage imposées par Google et regagner ainsi pleinement le domaine public.

Un procédé très astucieux

Ces Robins des Bois exploiteraient une faille des conditions d'utilisation de Google Books. Celles-ci interdisent en effet d'extraire du contenu de Google Books en procédant à des « requêtes automatisées » (type aspiration de base de données), mais rien n'interdit à une myriade d'individus de venir télécharger quelques ouvrages pour les déposer dans *Internet Archive*. Google n'a pas la possibilité de les attaquer, pas plus certainement qu'il ne peut agir contre Internet Archive, dans la mesure où le site ne fait qu'héberger des contenus chargés par ces usagers.



J'avais déjà eu vent de telles pratiques au détour d'un **billet de mars dernier écrit par Peter Hirtle** sur LibraryLaw Blog. Il s'interrogeait à propos de ces agissements : « *One has to wonder about the utility of well-meaning efforts such as those of the **Internet Archive** to capture and store copies of the Google scans* ». En cherchant bien dans les forums d'Internet Archive, on trouve effectivement des discussions relatives à des chargement de livres du domaine public extraits de Google Books (voyez **cet échange du 29 juillet 2010**).

Un watermark contesté

Ailleurs, on apprend également que des utilisateurs d'Internet Archive font disparaître le watermark « Numérisé par Google » que la firme appose au bas de toutes pages des ouvrages de sa bibliothèque numérique », car ils considèrent qu'une telle marque n'a aucune valeur juridique, lorsqu'elle est apposée sur un ouvrage du domaine public.



This would be illegal on Google's part since no one can copyright or restrict the text part of a public domain item [...] we can and should remove their watermarks, since once again this is after all public domain stuff. The founding fathers of the United States would turn-over in their graves is the actually saw what the political scum in this country have done to the copyright laws. Protection was never meant to be for all eternity or for some parasitic mega-corporate mass-media corporations.



Une fois « libérés » dans Internet Archive, les livres sont placés sous le statut « Public Domain » qui permet tout type de réutilisation, **y compris à des fins commerciales.**

Finalement, à la lumière de cet exemple, comme de celui que j'avais repéré mercredi, on se rend compte que les restrictions que Google tente d'imposer pour verrouiller son contenu sont peut-être beaucoup plus fragiles qu'on ne pourrait penser. Et ce qui va également dans ce sens, c'est que malgré ces extractions massives, Google n'a pas réagi pour l'instant.

Merci à Pablo d'avoir attiré mon attention sur ce point. Je copie ici in extenso son commentaire qui vous apportera des précisions supplémentaires :



J'ai réalisé il y a quelques semaines que plus de 900 000 livres avaient été extraits à partir de Google Book Search et archivés sur Internet Archive. Ce « détournement » des documents du domaine public, scannés par Google à partir des fonds des bibliothèques m'a bien plu car c'est un joli cas d'arroseur arrosé... le même « fair use » évoqué par Google peu alors être utilisé par un autre organisme comme l'internet archive pour se justifier.

Malgré le fait que les PDFs ne sont pas diffusés par l'archive (il y a toujours un lien sur la version de Google qui pourrait compenser l'abus) le full text est bel et bien proposé dans différents formats au téléchargement et à la consultation en ligne via son propre système.

Selon la courte explication donné par Internet Archive sur cette page :

« Digitized books from many different libraries from the Google Book Search program. These digital files have been downloaded from the Google site and uploaded to the Internet Archive by users. While these books may be old enough to be in the public domain, but there is no guarantee by anyone of their legal status. These books have been made text searchable as a finding aid and downloading refers to Google's site. Please refer to Google's site for any rights issues or restrictions. »

Ils ont trouvé la parade qui va dans le même sens que celui montré dans ce billet : si les documents ont été extraits par les utilisateurs, cela exclue l'extraction par des moyens automatiques tant redoutée par Google. Cela me paraît très intéressant car, au final, si les utilisateurs se mettent à télécharger en masse les PDF de Google des œuvres du domaine publique, à faire tourner l'OCR dessus et rediffuser sa propre version en texte brut rien ne pourrait les empêcher.

En regardant de près les 900 000 documents on voit très vite que c'est toujours le même utilisateur qui a fait le boulot, ce qui est encore plus surprenant ! Mais tout s'explique quand on lit son code d'utilisateur : « TPB » (les pirates sont encore parmi-nous, hip-hip !)

J'ai tenté de chercher des explications ou un semblant de partenariat du côté de Google ou ailleurs et je n'ai rien trouvé pour le moment, jusqu'à la parution de votre billet qui ajoute une nouvelle lumière et un trou inattendu pour lequel ces documents peuvent retourner dans le domaine public. Le plus drôle de l'histoire est que maintenant il y a des livres pour lesquels la version européenne de Google Books ne donne que des extraits mais qu'on peut télécharger complètement via Internet Archive car ils ont exploité le texte intégral fourni par la version américaine du moteur, par exemple cette version des travaux de Charles Darwin publié en 1896, qui dépasse la date fatidique de 1870 pour nous, pauvres européens .

Je me réjouis de voir la réaction de Google (si réaction il y en a...) mais je doute qu'ils tentent quelque chose contre ce détournement car ils attaqueraient le principe fondateur du fair-use qui équivaldrait à se tirer une balle dans le pied. Et finalement ce ne sont que 900'000 livres ...



>> Article publié initialement sur [S.I.Lex](#)

>> Illustration Flickr CC : [Jameson42](#)

TOTO

le 15 octobre 2010 - 0:05 • SIGNALER UN ABUS - PERMALINK



Je pense que c'est domaine public aux USA (avant 1923). En Europe, et dans beaucoup d'autres pays, tous les livres publiés après 1870 ne sont pas tous dans le domaine public.

VOUS AIMEZ



VOUS N'AIMEZ PAS



LUI RÉPONDRE

1 ping

A Google Books esete Robin Hoodal - Der Systemdenker le 26 octobre 2010 - 9:54

[...] – a tudományos diskurzust a digitális újságírás módszereivel fellendíteni szándékozó OWNIsciences cikkírója (pontosabban: blogolója) szerint – az Internet Archive (IA), illetve annak [...]