

Google a-t-il le droit de "scrap" nos contenus ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Google propose de plus en plus dans ses résultats des extraits de pages web, comme il l'a montré récemment en reprenant purement et simplement des tutoriels directement dans ses SERP sans demander l'avis des éditeurs les ayant créés sur leurs sites respectifs. A-t-il juridiquement le droit de mettre en place de telles pratiques ? Peut-il être attaqué pour "vol de contenu", ce à quoi peut être assimilé le "scraping" ? Réponses dans cet article...

Le post de Matt Cutts, en décembre 2013, sur le scraping et la mise à disposition d'un formulaire de signalement de ce type de cas (<http://www.abondance.com/actualites/20131205-13450-matt-cutts-et-le-scraping-de-textes.html>) a rouvert le débat autour de cette pratique. Et ce, d'autant plus étrangement que, concomitamment et de plus en plus, Google recopie quelques lignes de contenu de certains sites (comme Wikipédia) dans le cadre de la fourniture de résultats naturels (<http://www.abondance.com/actualites/20140626-14037-knowledge-graph-commence-afficher-tutoriels-scrape-web-vergogne.html>). La nature et les risques juridiques de cette pratique sont pourtant précis et ne souffrent pas de discussion. Les exemples de procès perdus par Google Actualités dans le cadre de la reprise de filets d'actualités, sont là pour en témoigner.

Le concept de « scraping »

Le terme « scraping » vient de l'anglais et peut signifier « racler » ou « prélever ». Ce terme est notamment utilisé dans un sens médical.

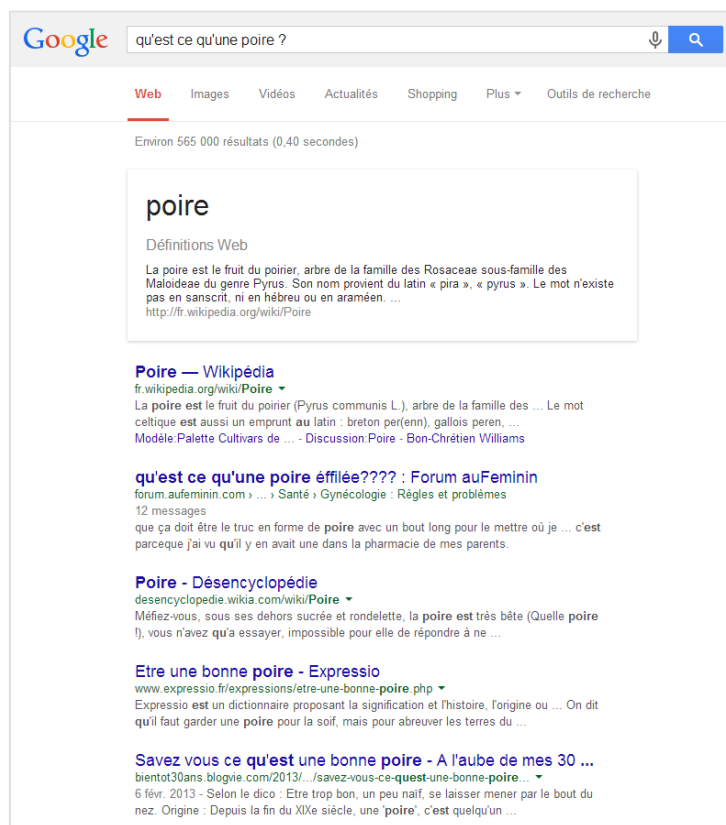
Il semble exister plusieurs cas de scraping :

- Selon Wikipedia (http://fr.wikipedia.org/wiki/Web_scraping), le *Web scraping* (parfois appelé *Harvesting*) est une technique d'extraction du contenu de sites Web, via un script ou un programme, dans le but de le transformer pour permettre son utilisation dans un autre contexte, par exemple le référencement ;
- Le scraping peut concerner uniquement les images ou vidéos, voire des écrans d'impression de sites ;
- Le scraping peut viser une technique d'aspiration de données structurées, dans une base de données, qui peut être réalisée entre deux applications (par exemple, en cas de clusterisation de données).

La notion induite est donc la copie de données.

Concrètement, le scraping peut viser tout ou partie d'un site.

- Il semble exister des pratiques consistant à recopier intégralement un site ;
- De nombreux sites ou moteurs de recherche reprennent des extraits de site (comme dans l'exemple ci-dessous avec Google reprenant les premières lignes du contenu de la page Wikipédia).



Google peut-il « scraper » le contenu de tiers ?

En faisant apparaître un filet de quelques lignes reprenant le contenu d'un site tiers, Google pratique donc clairement le scraping.

D'un point de vue juridique, cette reprise pourrait apparaître comme de la violation de droit d'auteur (ou copyright dans les pays anglo-saxons).

En pratique, le régime du droit d'auteur exige que l'œuvre copiée soit originale et une création de forme perceptible par les sens (la vue pour ce qui est des écrits). La condition d'originalité requiert que l'œuvre porte l'empreinte de la « personnalité » de son auteur. L'originalité est un concept distinct de celui de « nouveauté ». Ainsi, le contenu d'un site, un texte, une photo, etc... n'est protégé par le droit d'auteur que s'il est original dans le sens qu'il porte l'empreinte de la « personnalité » de son auteur. Les tribunaux font attention à ce principe directeur lorsqu'ils déterminent si oui ou non une œuvre est protégée. Ainsi, contrairement à ce que beaucoup de personnes pensent, un contenu banal qui n'a rien d'original n'est pas protégé par le droit d'auteur et l'auteur n'a pas le droit d'interdire à un tiers de le copier.

Si un contenu est protégé au titre du droit d'auteur, il n'en reste pas moins qu'un tiers a toutefois le droit d'utiliser une courte partie de ce contenu dans certaines conditions : l'article L122-5 du Code de la Propriété Intellectuelle dispose précisément que « *lorsque l'œuvre a été divulguée, l'auteur ne peut interdire (...), sous réserve que soient indiqués clairement le nom de l'auteur et la source les analyses et courtes citations justifiées par le caractère critique, polémique, pédagogique, scientifique ou d'information de l'œuvre à laquelle elles sont incorporées (...)* », étant précisé que cette exception ne peut « *porter atteinte à l'exploitation normale de l'œuvre ni causer un préjudice injustifié aux intérêts légitimes de l'auteur* ».

Il faut alors :

- Que le nom de l'auteur et la source soient mentionnés clairement ;
- Que l'extrait de l'œuvre originelle soit « incorporé » dans un ensemble plus large. Cette incorporation suppose que l'ensemble plus large ait un caractère « critique, polémique, pédagogique, scientifique ou d'information » qui suppose un travail d'analyse plus général où les extraits ont vocation à exposer le propos. ;
- Que la citation réponde à l'exigence de brièveté, n'être qu'un accessoire de l'œuvre dans laquelle elles s'insèrent : pour en juger, le critère traditionnellement reçu consiste à apprécier si l'œuvre dite seconde, celle qui recourt aux citations, peut survivre sans dénaturation à la suppression desdites citations. Cependant la loi ne prévoit pas de "taille mesurable" pour cette notion de "citation", qu'il est du coup complexe de quantifier en termes de mots ou de caractères mais simplement en "exigence de brièveté"... Ainsi, dans le passé, certains tribunaux ont accepté la reprise de 50% d'un passage, et d'autres courts ont condamné pour 10% repris...

Si la première condition est remplie correctement par Google (que ce soit dans l'exemple du résultat naturel ou dans le cadre de Google Actualités), la troisième condition peut parfois prêter à discussion, notamment quand le contenu visé est si court que la reprise consiste, en réalité, en une copie totale du contenu. Enfin, et surtout, la deuxième condition nous semble extrêmement discutable dans de nombreux exemples de services de Google. En effet, l'exemple du résultat naturel ne s'inscrit pas dans un cadre plus général et structuré. Au demeurant, Google Actualités a connu, sur le terrain de la courte citation, plusieurs batailles souvent perdues ou transigées. En effet, Google reprend quelques lignes de chaque article et agrège le tout. L'AFP avait ainsi ouvert le bal des procès pour la reprise des titres, quelques lignes et photos, la procédure se terminant par une transaction évidemment secrète. Toutefois, dans d'autres pays de droits d'auteur, comme la Belgique ou le Brésil, Google a connu des déconvenues et a été contraint de désindexer les articles de nombreux journaux nationaux. L'argument de la courte citation n'a manifestement pas été reçu par les tribunaux.

En conséquence, Google ne peut pas juridiquement reprendre des contenus de tiers de manière systématique et automatisée sans l'accord des auteurs sur le terrain de la courte citation.

Si le droit d'auteur est très vaste et permet à un juge de dire parfois oui et parfois non à la protection d'une œuvre, il existe toutefois une autre voie : celle du parasitisme. Trois conditions sont nécessaires pour se voir reconnaître responsable d'un fait :

- Faire une faute ;
- Qu'il y ait un préjudice pour quelqu'un ;
- Qu'il y ait un lien de causalité entre la faute et le préjudice.

Ainsi, si la reproduction d'une œuvre n'est pas une contrefaçon, elle peut constituer parfois une faute. Certains se souviendront que les Conditions Générales de Vente du site VentePrivee.com avaient été copiées et utilisées par un concurrent. Ce concurrent a été condamné à 30.000 € pour parasitisme : la Cour d'appel de Paris a estimé que les Conditions Générales de Vente ne sont pas protégées par le droit d'auteur, mais que le fait de les copier est un acte de parasitisme condamnable.

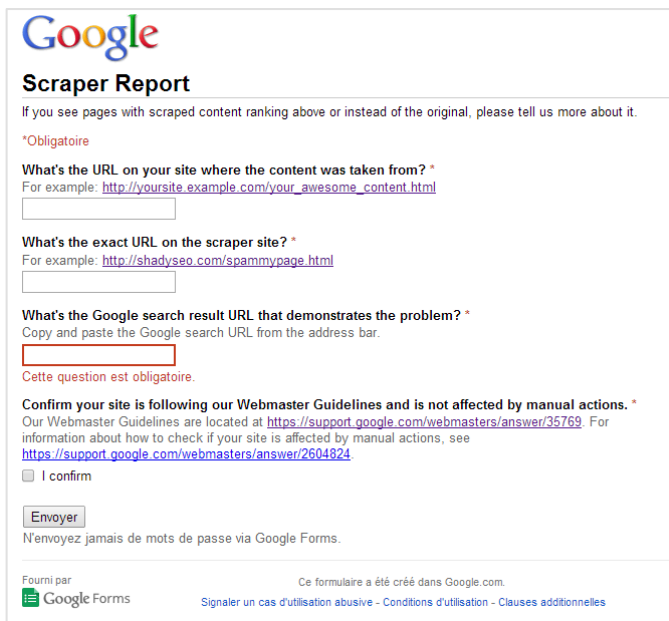
Des internautes peuvent-ils « scraper » le contenu de tiers ?

Par ailleurs, tout un chacun peut parfaitement être concerné par la question : soit votre site a été « scrapé », soit vous « scrapez » le contenu d'un tiers.

Les règles juridiques sont exactement les mêmes que pour Google : le droit d'auteur protège un contenu original et le copier est pénalement et/ou civilement répréhensible. L'exception de courte citation peut permettre de copier de courts extraits, mais dans le cadre exposé ci-avant.

Les internautes doivent, eux, faire attention à un point supplémentaire : les règlements de Google. En effet, Matt Cutts a annoncé, en février 2014, la mise en place d'un formulaire « Scraper Report »

(https://docs.google.com/a/google.com/forms/d/1Pw1KVOVRyr4a7ezj_6SHghnX1Y6bp1S0Vmy60QjkF0Y/viewform) permettant de signaler à Google des sites qui ont été copiés.



The image shows a screenshot of a Google Forms page titled "Scrapper Report". At the top is the Google logo. Below it, the title "Scrapper Report" is followed by a brief instruction: "If you see pages with scraped content ranking above or instead of the original, please tell us more about it." A red asterisk indicates a required field. The first question is "What's the URL on your site where the content was taken from?*" with an example URL and a text input field. The second question is "What's the exact URL on the scraper site?*" with an example URL and a text input field. The third question is "What's the Google search result URL that demonstrates the problem?*" with a red-bordered text input field and a note "Cette question est obligatoire." The fourth question is "Confirm your site is following our Webmaster Guidelines and is not affected by manual actions.*" with a checkbox labeled "I confirm" and a link to Google's Webmaster Guidelines. At the bottom, there is an "Envoyer" button, a note "N'envoyez jamais de mots de passe via Google Forms.", and footer information including "Fourni par Google Forms" and "Ce formulaire a été créé dans Google.com." with links for reporting abuse, terms of use, and additional clauses.

Rien n'interdit de plus, de signaler la copie du site par le biais du formulaire DMCA de Google (<https://support.google.com/legal/troubleshooter/1114905?rd=2-ts=1115655,1282900,1115846,1117010,1697925>) dans la mesure où le scraping porte souvent sur du contenu protégé.

Enfin, et de manière plus large, reste la question de savoir si le scraping est un acte de « negative SEO ». En effet, la copie servile de site tiers peut avoir pour fonction de « pousser » certaines pages en référencement. Nous estimons que la réponse ne peut être que positive car, tant dans l'esprit que dans les manipulations techniques, le scraping peut viser à optimiser de manière illicite le référencement. Au demeurant, la rubrique « Contenu de Qualité » des « Conseils pour Webmasters » est claire à ce titre : le scraping n'apporte rien, peut porter atteinte à des droits de tiers et induit que le contenu du site servile ne sera pas de qualité (et donc répréhensible dans le langage de Google) (<https://support.google.com/webmasters/answer/2721312?hl=en>).

En conséquence, en plus de risquer d'éventuels problèmes juridiques, les internautes scrapant un site pour le copier dans un autre, risquent une sanction de la part de Google en termes de référencement.

Alexandre Diehl
Avocat à la Cour, cabinet Lawint (<http://www.lawint.com/>)