

*Spécialiste du Web depuis plus de quinze ans, **Dave Hamel** est Chef de produits pour l'analytique Web, le référencement et le moteur de recherche Google à CBC/Radio-Canada. Il a travaillé à des sites pour le compte de clients tels que Citibank, Rolex et le réseau Thankyou. Dave est passionné de vélo et membre du club cycliste Lapdogs.*

Il n'y a pas si longtemps, pour trouver un renseignement, il fallait se rendre dans une bibliothèque ou consulter des encyclopédies couvertes de poussière. C'était avant Internet, qui nous offre un accès instantané à l'information. Mais il subsistait un problème : comment trouver un renseignement dans cet océan de données? Au tout début, il fallait consulter des répertoires semblables aux pages jaunes d'un bottin téléphonique, notamment Yahoo, Excite et DMOZ, qui analysaient les sites Web et les classaient dans différentes catégories. Or, les spécialistes du marketing n'ont pas tardé à abuser de ces répertoires dans le but d'accroître la fréquentation de leurs sites, de sorte que ces outils ont perdu de leur valeur et de leur utilité.

Puis, Google a changé les règles du jeu. Ce moteur de recherche, qui exécute plus de 65 % des recherches en anglais (*source : rapport qSearch de comScore, février 2011*), classe lui-même les sites Web au lieu de les laisser eux-mêmes désigner la catégorie à laquelle ils appartiennent. Ce procédé est connu sous le nom de recherche « organique », « naturelle » ou « algorithmique ». Il a transformé les méthodes de collecte de l'information et soulevé de nombreuses questions, notamment :

Comment Google décide-t-il quel site apparaît au premier rang des résultats?

Pourquoi mon site Web n'apparaît-il pas dans les résultats?

Comment puis-je améliorer ma position dans les résultats de recherche?

Heureusement, Google lui-même donne la réponse : il s'agit du processus de **référencement**, qui consiste à modifier un site Web de manière à rehausser son rang dans la page des résultats de recherche. Le processus ne relève pas uniquement des TI; il peut comprendre une part d'encodage, mais aussi la révision du texte ou du titre, la modification de l'ordre hiérarchique en vue d'améliorer la navigation de l'utilisateur, ou le renforcement du marketing et des relations publiques dans le site Web. Si vous êtes responsable d'un site Web, le référencement est votre principale préoccupation et se résume à une simple règle : **créer du contenu qui mérite d'être partagé.**

Comment connaissez-vous le meilleur restaurant en ville? Pourquoi faites-vous réparer votre voiture à tel garage? Pourquoi allez-vous voir un film indépendant plutôt que la dernière production hollywoodienne? Il est fort probable que votre choix ait été en partie influencé par ce qu'on vous a conseillé. Le référencement s'apparente au « bouche à oreille » et joue un rôle essentiel sur Internet, compte tenu qu'on recense plus de 270 millions de sites Web uniques et plus de deux milliards d'internautes, et que Google

exécute à lui seul plus d'un milliard de recherches par jour! Il importe donc de faire en sorte que les utilisateurs trouvent votre site.

En outre, parmi les résultats d'une recherche, le premier récolte 40 % des clics des utilisateurs, le deuxième 10 %, le troisième n'en obtient que 8,5 %, et le pourcentage de clics décroît rapidement pour les résultats suivants; par conséquent, si votre site Web ne figure pas sur la première page de résultats, les probabilités que les internautes y accèdent sont faibles. Or, cette courbe statistique présente une longue traîne. En effet, plus de 70 % des interrogations contiennent des mots qui font l'objet de cinq recherches ou moins par mois. Ainsi, il n'est pas nécessaire de se classer en tête de liste des résultats de recherches génériques. Par exemple, vous n'obtiendrez peut-être pas le premier rang pour le mot « haricots », mais peut-être pour le terme « haricots de Lima ».

Ce qui soulève donc une autre question : « Comment cela fonctionne-t-il? »

Google utilise un robot Web qui balaie Internet et collecte l'information. Les données recueillies sont ensuite indexées, et un algorithme y est appliqué afin de déterminer l'importance de la page. En fonction de nombreux facteurs, l'algorithme attribue l'indice PageRank (nom donné en l'honneur du fondateur de Google, Larry Page). Le PageRank est la représentation numérique de la valeur perçue d'une page Web par rapport aux autres pages contenues dans Internet, et comporte divers éléments, notamment :

- l'autorité du nom de domaine;
- la popularité des liens pointant vers la page;
- le texte d'ancrage des liens externes;
- l'utilisation des mots-clés sur la page;
- les données d'enregistrement et d'hébergement;
- le trafic et le taux de clics; et
- la mesure des médias sociaux.

L'autorité du nom de domaine désigne simplement la fiabilité ou la pertinence d'un site comparativement aux autres sites. En somme, il s'agit d'évaluer son importance pour les utilisateurs finaux. Par exemple, quel site vous paraît le plus sérieux : www.apple.com ou www.win-ipad-gamble.tv? Internet a décidé pour vous et accorde une plus grande importance au domaine Apple. Cependant, l'autorité du nom de domaine est un facteur complexe, dont la valeur est calculée en fonction de plus de 150 éléments. Il est donc très difficile de l'influencer directement.

La popularité des liens représente la somme des liens pointant vers une même page. On peut comparer ces liens à des votes en faveur de la page Web : plus il y en a, mieux c'est. Le texte d'ancrage de ces liens pèse aussi dans la balance. « Le gouvernement de l'Ontario ignore si les compteurs intelligents fonctionnent » est un exemple de lien très descriptif. Il donne à l'utilisateur un aperçu du contenu qui y est associé. À l'inverse, « cliquez ici » n'est pas du tout descriptif et peut être lié à n'importe quel type de contenu. Google accorde donc plus de poids au premier lien.

Si divers liens affichant un texte similaire pointent vers une page, Google lui accorde du crédit. Ainsi, pour reprendre l'exemple précédent, si un autre lien indique « Les compteurs intelligents sont défectueux, déclare le gouvernement de l'Ontario », l'algorithme de Google reconnaît la page comme une référence en ce qui concerne les compteurs intelligents et le gouvernement de l'Ontario. Pour cela, il suit le lien jusqu'à sa destination et lit le contenu de la page.

Il examine plus particulièrement la **balise titre**, qui sert de texte pour les signets, l'onglet du navigateur et le lien de la page dans Google. Il cherche également une balise *h1*, ou balise *en-tête*, comme le titre d'un article et le corps du texte.

Ce qui nous amène à l'aspect suivant de PageRank, à savoir la **densité de mots-clés** ou l'**utilisation des mots-clés dans la page**. Le robot Web a déjà relevé les mots-clés du lien; dans notre exemple, « compteur intelligent », « gouvernement » et « Ontario ». Il lit ensuite le texte de la page afin de vérifier que ces mots s'y trouvent et que la page traite bel et bien du sujet annoncé par le texte du lien.

Si le lien renvoyait à une page sur les singes et non sur les compteurs intelligents, Google n'en tiendrait pas compte. S'il trouvait plusieurs de ces liens, il soupçonnerait une escroquerie et commencerait à pénaliser la page. L'algorithme prend aussi en considération la lisibilité. En d'autres termes, inutile de semer le texte d'une multitude de « compteurs intelligents ». Votre contenu doit rester compréhensible pour les utilisateurs finaux.

Les données d'enregistrement et d'hébergement influent aussi sur le référencement. Google accorde plus de poids à un domaine « *.com* » qu'à un domaine associé à un pays tel que « *.ca* ». De plus, le domaine « *.com* » étant plus ancien, il y a de fortes chances que les noms de domaine composés d'un seul mot suivi de « *.com* » existent aussi depuis longtemps – ce qui leur vaut un rang élevé. L'algorithme tient compte de divers facteurs : le nombre d'années où votre site a été hébergé dans le même domaine, le nombre d'autres sites appartenant à ce domaine et les autres domaines associés à votre entreprise.

Les URL jouent aussi un rôle dans le référencement, quoique mineur. Une URL telle que www.cbc.ca/news/ontario-government-smart-meters-useless a plus de valeur que www.cbc.ca/n/?postid=536. Elle donne de l'information sur le contenu, et Google récompense cet effort.

Le trafic et le taux de clics sont les critères sur lesquels s'appuie Google pour vérifier ses propres résultats en fonction de ce que les utilisateurs cherchent. Par exemple, si la recherche « compteurs intelligents » produit deux résultats ayant un indice PageRank comparable, mais qu'un plus grand nombre d'utilisateurs a cliqué sur le deuxième lien, Google classera celui-ci au premier rang, car il considère le taux de réponse plus élevé comme un indicateur de la pertinence du résultat.

Enfin, avec l'essor des **médias sociaux** tels que Facebook et Twitter, les liens ne proviennent plus toujours de sites Web en soi, et ils ne s'inscrivent pas forcément dans un texte de forte densité. Or, les liens compris dans les médias sociaux sont importants de par leur large portée. L'algorithme de Google tient donc compte de cet aspect, bien que son importance demeure faible.

Le poids ou l'importance de chacune de ces valeurs évolue, car Google modifie constamment son algorithme. Personne ne peut dire avec certitude combien de liens dans Twitter équivalent à un lien dans un site Web, mais on sait toutefois que les liens provenant de sites respectés ou dotés d'un indice PageRank élevé valent plus que les liens issus de sites dont l'indice est faible. En clair, un lien de CBC/Radio-Canada pèse davantage qu'un lien de Davesblog.blogspot.com. Certains ont recours aux techniques dites « grises », comme les « fermes de liens » ou les pages passerelles, en vue d'influencer Google en leur faveur. Or, les techniques qui fonctionnaient hier ne sont d'aucune utilité aujourd'hui et pourraient bien être sanctionnées demain. Par conséquent, toute tentative de tromper Google risque de miner vos efforts à long terme.

Maintenant que vous savez ce qui importe aux yeux de Google, voici quelques règles qui vous aideront à rehausser votre rang dans les résultats de recherche.

Règle n° 1 – Établissez d'abord une base solide : ne vous préoccupez pas du marketing viral ou des partenariats susceptibles de vous valoir des hyperliens avant d'avoir créé un contenu de qualité accessible et unique, assorti de mots-clés spécifiques et ciblés. Autrement, votre site reposera sur une assise instable. N'oubliez pas que la recherche est conçue pour les internautes, et non pour les sites Web. Pensez d'abord aux utilisateurs finaux au moment de construire votre site.

Règle n° 2 – Du contenu de qualité, rien d'autre : ne créez que du contenu pertinent pour les utilisateurs, qui a pour **seules** fonctions de résoudre les problèmes, de répondre aux questions et d'informer. Demandez-vous : pourquoi un tiers créerait-il un lien vers mon contenu? Pourquoi souhaiterait-il le partager? Prenons par exemple cette page Web (<http://www.cbc.ca/stevenandchris/2011/07/seven-countertop-wine-racks.html>). Que pourriez-vous écrire à son sujet pour inciter les internautes à y jeter un œil? Quelle solution ou quelle information offre-t-elle?

Règle n° 3 – Efforcez-vous d’indiquer quatre à six mots-clés pour chaque page : le contenu se classe mieux dans les résultats de recherche s’il est ciblé. Consultez les dernières mesures Web afin de connaître les mots fréquemment employés pour trouver votre site. Ainsi, si vos clients tapent « souliers de course » pour vous trouver, n’utilisez pas l’étiquette « espadrilles », mais renforcez plutôt les mots-clés « souliers de course ».

Règle n° 4 – Les moteurs de recherche ne « voient » pas les images; il est donc essentiel d’y associer des attributs *alt*. En plus d’améliorer le référencement, ces attributs rendent votre site plus accessible. Veillez également à bien former votre code et à valider le W3C (<http://validator.w3.org/>). Ces détails consomment peu d’efforts et indiquent à Google votre réel souci des utilisateurs.

Règle n° 5 – Soignez votre balise *titre* : c’est le texte utilisé pour le lien dans les résultats de recherche sur Google, la barre titre dans le navigateur et le texte du signet; ne sous-estimez pas son importance. Par exemple, si votre page porte sur les haricots de Lima, rédigez un titre comme celui-ci : « Les haricots de Lima, l’aliment nutritif des Incas d’Amérique du Sud ». Ainsi, vous générez du trafic pour les termes « haricots », « Lima », « aliment », « nutritif », « Incas » et « Amérique du Sud ».

Règle n° 6 – Rédigez une méta-description informative : Google utilise ces balises pour la description du lien. Elles constituent votre meilleur argument pour attirer les internautes sur votre site. Annoncez-leur ce qu’ils y trouveront.

Règle n° 7 – Respectez le cycle du référencement : les internautes lisent votre contenu et, s’il est intéressant, ils créeront un hyperlien. À mesure que les liens pointant vers votre site se multiplient, son classement dans les résultats du moteur de recherche s’améliore. Par conséquent, les internautes sont plus nombreux à consulter votre contenu de qualité. **Et le cycle se répète, et votre site gagne en importance.**

The screenshot shows a web browser displaying a CBC News article. Red boxes highlight the following elements:

- The browser's address bar containing the URL: `www.cbc.ca/news/technology/story/2012/01/05/tuna-tokyo-record.html`.
- The main headline: **Giant tuna caught in Tokyo fetches record \$747K**.
- The sub-headline: **Winning bidder at auction heads sushi restaurant chain**.
- The main image showing sushi chefs with a large piece of bluefin tuna.
- A video player with the title "\$747K tuna" and a duration of 0:53.
- Text in the video player: "A bluefin tuna caught off northeastern Japan fetched a record \$6.49 million yen (about \$746,797 Cdn) on Thursday in the first auction of the year at Tokyo's Tsukiji fish market."
- Navigation tabs for "Technology & Science" and "Community".
- Advertisement for "REDEMPTION INC" with the text "MONDAYS AT 9PM/9:30PT SERIES PREMIERE JANUARY 9TH".
- Top News Headlines section featuring a photo of Barack Obama and the text "Obama unveils leaner U.S. defence strategy".

CBC/Radio-Canada observe généralement ces règles, bien qu'il y ait matière à amélioration. L'article reproduit ci-dessus est intitulé « Giant tuna caught in Tokyo fetches record \$747K ». Les mots-clés sont « giant », « tuna », « caught », « Tokyo », « fetches » et « record »; on s'attend donc à retrouver ces mots-clés dans la page. Commençons par l'URL : il contient « tuna-tokyo-record.html », soit seulement 50 % des mots-clés. Il y a lieu de l'améliorer.

La balise *titre*, représentée par le titre de l'onglet, en haut à gauche, celui du signet et le texte en surbrillance, contient tous les mots-clés, suivis de l'information hiérarchique, la section et le site Web. C'est un bon point.

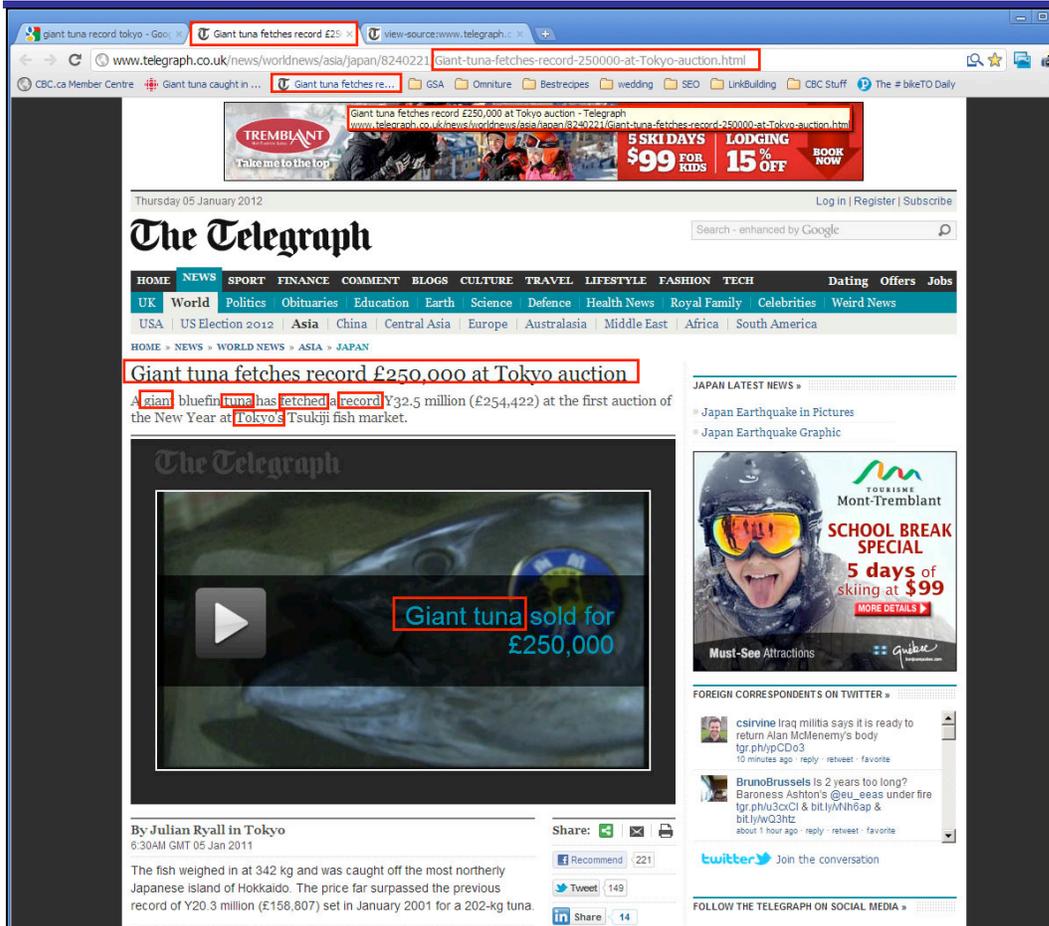
Le début de l'article contient les mots « tuna », « caught », « fetch », « Tokyo » et « record », mais il manque le mot « giant ». L'emploi de tous les mots-clés renforce l'idée maîtresse de la page, soit : « a giant tuna fish that was caught and fetched a record price in Tokyo ». *Vous voyez la différence?*

L'examen du résultat de recherche dans Google fournit d'autres indications sur les moyens d'améliorer le classement et d'optimiser le site Web.

[Tuna caught in Tokyo fetches record \\$747K - CBC News](http://www.cbc.ca/news/offbeat/story/2012/01/05/tuna-tokyo-record.html)
www.cbc.ca/news/offbeat/story/2012/01/05/tuna-tokyo-record.html
29 minutes ago – A bluefin **tuna caught** off northeastern Japan **fetches** a record 56.49 million ... Skip to main content; Skip to **Tuna caught in Tokyo fetches** record \$747K ... U.S. election 2012 Iowa the **big** loser; what's in store for Romney CBC ...

Première observation : l'URL dans Google diffère de celle que nous avons vue plus haut. Cette dernière était classée sous « technology », alors que dans Google, elle se classe sous « offbeat », ce qui pose un problème. En effet, lorsque Google trouve du contenu en double, il indexe l'une des versions et élimine l'autre résultat. Par conséquent, si des blogueurs créent un lien pointant vers la version « technology » de l'article au lieu de la version « offbeat », Google ne tient pas compte de ces liens, puisqu'il a éliminé la version « technology ».

De plus, la description contient un passage qui n'a rien à voir avec l'article, c'est-à-dire la dernière phrase sur les élections aux États-Unis. L'examen du code indique un ajout délibéré à des fins d'accessibilité; or, il apparaît au mauvais endroit.



Lorsqu'on cherche « giant tuna fetches record Tokyo », le premier résultat renvoie à une page du journal britannique *The Telegraph*. Cela s'explique en partie par le nom de la page dans l'URL, qui contient tous les mots-clés suivants : Giant-tuna-fetches-record-250000-at-Tokyo-auction.html. En outre, le site se classe au premier rang parce que la première phrase de l'article contient un certain nombre de mots-clés.

Enfin, la méta-description de la page du *Telegraph* suit immédiatement la balise *en-tête*, après la balise *titre*, ce qui contribue à améliorer le référencement.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en"
xmlns:fb="http://www.facebook.com/2008/fbml" xmlns:og="http://opengraphprotocol.org/schema/">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />

<title>Giant tuna fetches record £250,000 at Tokyo auction - Telegraph</title>

<meta name="description" content="A giant bluefin tuna has fetched a record Y32.5 million
(&#pound;254,422) at the first auction of the New Year at Tokyo's Tsukiji fish market." />

<meta name="keywords" content="tuna, giant, tokyo, japan, auction, sushi, Japan,Asia,World
News,News" />
```

Le nombre de liens retour vers le domaine, qui représente le nombre de « votes » qu’obtient le site, constitue un facteur clé, bien qu’on ne puisse pas le voir. *The Telegraph* compte environ 1 110 000 liens retour, contre seulement 638 000 pour CBC/Radio-Canada. La page du quotidien britannique fait l’objet de 149 gazouillis, et celle de CBC/Radio-Canada, trois. Souvenez-vous que le bouche à oreille est la clé du succès, et dans l’univers du Web, les médias sociaux en sont les vecteurs.

Notons quelques problèmes techniques : par exemple, CBC/Radio-Canada n’offre pas de plan de son site au format XML qui permet de guider le moteur de recherche vers les pages. Il ne s’agit pas d’un élément essentiel, mais la moindre amélioration constitue un atout.

À la lumière de ces observations, avec quelques changements mineurs et un peu de bonne volonté, nous pouvons améliorer notre classement dans les résultats. Notons enfin que certaines des modifications nécessaires touchent le code, d’autres la hiérarchie, et d’autres encore concernent le texte.

L’art du référencement ne relève pas de la magie noire, ni même de la magie. Avec un peu de patience, il permet d’améliorer progressivement le classement du site dans les résultats, grâce à un texte enrichi, à des titres plus descriptifs, à un code et à une architecture plus nets, à des mots-clés ciblés et à une planification réfléchie.

Google met tout en œuvre pour que les recherches produisent des résultats pertinents pour les utilisateurs finaux. Si vous suivez son exemple, il récompensera vos efforts et, bientôt, vous vous hisserez au premier rang des résultats de recherche.