

# Google, meilleur ennemi de la presse en ligne

written by Guillaume Sire | 9 septembre 2014

Hormis quelques dizaines d'employés chez Google, personne ne sait exactement quels sont les critères qui président à la hiérarchisation des informations sur le moteur de recherche. Cependant, il existe des méthodes permettant aux éditeurs de contenus d'augmenter leurs chances de figurer en haut des classements. De plus en plus d'écoles de journalisme proposent des cours d'« optimisation » pour Google – couramment appelée *Search Engine Optimization* ou SEO –, dont l'objectif est d'apprendre aux journalistes à utiliser ces méthodes afin de maximiser la visibilité de leur production sur le web. Cependant, la SEO soulève une question importante : n'est-elle pas susceptible, en effet, d'avoir une influence sur l'aspect éditorial de la production journalistique et, donc, potentiellement, sur le traitement de l'actualité et sur l'actualité elle-même, dans la mesure où l'on considère à la suite de Pierre Nora que le traitement médiatique n'agit pas indépendamment des événements dont il rend compte, mais constitue au contraire « la condition même de leur existence »<sup>1</sup> ?

Lorsqu'ils suivent l'une de ces formations appelées « écriture web », les journalistes se rendent compte qu'ils devront faire certaines concessions afin de figurer en tête des classements produits par Google. En effet, lorsqu'on sait que, pour attirer un trafic substantiel, il faut traiter de préférence tel ou tel sujet, et utiliser tel ou tel mot, la question se pose de savoir à quel point l'offre (le travail des journalistes) s'adaptera à la demande (les sujets en vogue sur Google ainsi que les préconisations données par les concepteurs du moteur) et à quel point les journalistes devront « revoir » la vision du monde qu'ils souhaiteraient idéalement communiquer. Alors que les chercheurs se sont encore peu intéressés à ce débat, de nombreux journalistes s'interrogent sur la place qu'occupent dans leur pratique professionnelle les enjeux liés au référencement. Ils se demandent si, pour être lus par les utilisateurs des services de Google, ils pourront écrire ce qu'ils souhaitent et comme ils le souhaitent, ou bien s'ils devront au contraire accepter de faire des concessions quant à leur déontologie. Autrement dit, ils se posent la question de savoir comment leur crédibilité et leur visibilité, ainsi que la rentabilité de l'entreprise qui les emploie, pourront aller de pair, dès lors qu'ils cherchent à ce que l'information qu'ils produisent s'inscrive à l'avant-scène de l'espace numérique. Cette question doit être considérée dans un contexte de crise où les entreprises de presse peinent à équilibrer leurs comptes et où les journalistes sont contraints d'osciller entre une logique industrielle et une logique citoyenne ; écartelés, comme dit le

sociologue Rémy Rieffel, « entre l'obligation de respecter des impératifs de vente, d'audience, et leur souci d'assurer au mieux l'autonomie de leur pensée »<sup>2</sup>. L'objectif ici est de rendre compte de la tension qui se crée entre les impératifs liés au référencement sur Google et les représentations des journalistes, ainsi que des implications concrètes de cette mise en tension en matière de « positionnement » du contenu sur la Toile, d'organisation du travail dans les entreprises de presse et de traitement de l'actualité.

### **Le PageRank Sculpting et l'égocentrisme hypertexte**

Le critère de pertinence le plus connu grâce auquel Google hiérarchise les contenus sur le web est sans aucun doute le *PageRank*. Cet indicateur a été mis au point par Sergey Brin et Larry Page, les deux fondateurs de Google, en 1998, année de création du moteur. Parce qu'il a fait l'objet de deux publications scientifiques<sup>3</sup> et de deux brevets américains<sup>4</sup>, le *PageRank* est celui des indicateurs de pertinence que l'on connaît le mieux. Il est cependant critiqué en raison de sa propension à permettre aux documents déjà visibles sur le web de devenir encore plus visibles sur Google, et à nourrir ainsi un cercle vicieux où les riches deviendraient encore plus riches<sup>5</sup>.

Schématiquement, le *PageRank* consiste à considérer que chaque lien hypertexte est un vote pour la page vers laquelle il pointe. Plus une page reçoit de votes, plus elle est considérée comme pertinente, et plus son vote, lorsqu'elle pointe elle-même vers d'autres pages, est revêtu d'autorité. Cela explique pourquoi la méthode de SEO la plus usuelle – appelée *PageRank Sculpting* – consiste à effectuer des liens hypertextes vers soi-même plutôt que vers des contenus extérieurs. Ainsi, l'autorité d'une page est redistribuée par son auteur aux autres pages du même site. Cela peut encourager de la part des éditeurs de contenus un certain repli sur soi, qui interpelle les journalistes dans la mesure où ces derniers ne sont pas encouragés à citer leurs sources ou leurs confrères œuvrant pour des titres concurrents.

Lorsqu'on observe le maillage des sites de presse en ligne, on constate un certain « égocentrisme hypertexte » dû aux stratégies du *PageRank Sculpting*. Pour ne pas peser sur les épaules du journaliste, la production de liens pointant vers l'intérieur peut être automatisée. Une liste de mots-clés est alors définie en fonction de l'actualité (Sotchi, Obama, Hollande...), et l'utilisation par un journaliste dans le corps de son article d'un de ces mots générera automatiquement un lien vers la « page d'atterrissage » associée. Sur cette page, l'ensemble des articles traitant de ce sujet seront répertoriés sous la forme d'une liste de liens ordonnés par ordre antéchronologique. Ainsi, l'autorité *PageRank* est systématiquement renvoyée depuis des articles à valeur ajoutée vers des pages jouant le rôle de fiches d'indexation et positionnées en tant que telles sur le moteur de recherche Google. Dès lors, les journalistes

peuvent se concentrer sur la production de liens à valeur ajoutée, pointant vers des articles publiés sur le même site ou vers des sources extérieures. Néanmoins, il se trouve que si les journalistes éditent trop de liens manuellement, ceux-ci diluent l'autorité distribuée par les liens pointant automatiquement vers les pages d'atterrissage. En effet, le *PageRank* fonctionne de telle façon que lorsqu'une page disposant d'une autorité [x] effectue [n] liens vers d'autres pages, chacun de ces liens aura une valeur équivalente à  $[x/(a.n)]$ , où [a] varie selon que le lien est effectué vers l'extérieur ou vers l'intérieur (les liens effectués vers l'extérieur peuvent ainsi être davantage valorisés par Google que les liens effectués vers l'intérieur). Dès lors, un usage trop intensif de la part des journalistes du lien hypertexte pourrait nuire à la stratégie de SEO. C'est pourquoi un dispositif d'alerte peut être mis en place visant à limiter l'action des journalistes, sans pour autant que ceux-ci soient destitués de leur capacité à effectuer des liens hypertextes : une alerte les prévient simplement lorsqu'ils ont dépassé le « quota idéal » de liens. Dans certains cas extrêmes (qui n'existent pas en France), ce dispositif peut être transformé afin d'empêcher la validation des articles dont les auteurs auraient dépassé le quota de liens autorisés.

En général, les liens dont le nombre est limité sont répartis de la manière suivante : un tiers pour le référencement et deux tiers pour les journalistes. Parce que ces liens n'ont pas tous le même but, les premiers pourront être invisibles, tandis que les seconds seront chargés d'une fonction de monstration et feront véritablement partie du traitement de l'actualité. En cas de conflit, ce sont les liens SEO qui cèdent leur place aux liens journalistiques. Dans une telle situation, la fonction journalistique attribuée au lien prime donc sur la fonction d'optimisation.

Il apparaît ici clairement qu'une tension existe entre les considérations liées à la visibilité du contenu et les considérations purement journalistiques, donnant lieu à des arbitrages et à des ajustements de la part des journalistes. Les stratégies du *PageRank Sculpting* induisent finalement un repli sur soi de la part des éditeurs soucieux d'optimiser leur référencement, et une dilution de l'autorité des liens effectués manuellement par les journalistes.

### **Le référenceur dans les salles de rédaction**

Les entreprises de presse sont nombreuses à embaucher des salariés chargés de la stratégie de référencement, appelés en général « référenceurs » ou « chargés de projet SEO ». C'est le cas par exemple en Angleterre de BBC News, *The Guardian* et *Northcliffe* comme l'a expliqué le sociologue Murray Dick , et c'est le cas en France dans les principaux sites de presse généralistes (Lemonde.fr, Lefigaro.fr, Nouvelobs.com, 20minutes.fr, Lepoint.fr, Lexpress.fr). Dans ces

entreprises, la stratégie de SEO n'est jamais « dogmatique ». Des négociations sont menées par le référenceur avec les différents services concernés, et si des pressions peuvent être exercées sur le journaliste, elles ne donnent pas lieu pour autant à des obligations strictes ou à des sanctions. C'est un travail de moyen et long termes qu'effectue le référenceur, qui vise à convaincre les journalistes du bien-fondé de la rationalisation des stratégies de référencement, et à concilier (ou réconcilier) les exigences de Google et la déontologie des journalistes.

Le champ d'action des référenceurs est triple : éditorial, technique et commercial. Pour ce qui est de la partie éditoriale, il peut arriver que les référenceurs participent aux conférences de rédaction. Leur fonction est alors d'informer les journalistes des sujets qui sont les plus recherchés par les internautes sur Google et qui, par conséquent, pourraient attirer de nombreux visiteurs si une page leur était consacrée. Les journalistes choisissent alors de traiter ou non ces sujets, en fonction de leur ligne éditoriale et de leurs propres représentations. Il arrive qu'il y ait des confrontations entre les référenceurs et les journalistes, mais ce sont toujours les journalistes qui ont le dernier mot. L'absence de rapport hiérarchique entre le référenceur et les journalistes est en effet un élément fondamental des stratégies de SEO.

De l'action du référenceur dépendent en fait deux choses distinctes : la maximisation du trafic, certes, mais également le contrôle exercé par l'entreprise de presse sur les modalités de publication et de diffusion de ses contenus. Le référenceur peut en effet utiliser certains outils visant à signaler à Google quelles sont les pages à indexer en priorité et comment elles devront apparaître. Ainsi, le SEO ne sert pas seulement à attirer de nouveaux lecteurs, mais également à avoir la main sur la manière dont le contenu sera présenté sur les listes de résultats générés par le moteur : quels sont les mots, les images, les extraits et les noms d'auteur qui apparaîtront, sous quelle forme et à quelles conditions. Par ailleurs, la stratégie du référenceur a aussi pour but de dissimuler certains des efforts qui sont faits dans le seul but d'attirer de l'audience et qui n'ont pas forcément de sens du point de vue journalistique. Ainsi, si son action réussit parfaitement, l'internaute ne se doutera pas de la présence de certains éléments destinés à optimiser le référencement, car ceux qu'il verra auront un intérêt, et un sens, à ses yeux, tandis que les autres ne lui seront jamais montrés (par exemple, les liens vers les pages d'atterrissage mentionnés ci-avant peuvent être invisibles).

### **Quel genre de journalisme Google récompense-t-il ?**

Plusieurs chercheurs en sciences sociales ont montré qu'un petit nombre de sites et un petit nombre de sujets d'actualité étaient mis en avant sur Google<sup>7</sup>. Selon ce que certains d'entre eux nomment la *Googlearchy*<sup>8</sup>,

la diversité des sources et le pluralisme des sujets traités sur le web – qui sont quasiment infinis – sont en effet entravés par le moteur de recherche qui concentre l'attention des internautes sur quelques sources et quelques sujets. Or, dès lors que les référenceurs assistent aux conférences de rédaction et que le fait qu'un sujet ait un fort potentiel sur Google puisse être prescripteur en matière de ligne éditoriale, la question se pose de savoir si Google n'a pas un effet négatif sur le pluralisme et la diversité de l'information. En effet, l'absence de diversité chez Google pourrait provoquer une absence de diversité chez ceux qui ont besoin de passer par le moteur pour accéder à leur lectorat. Quant à la diversité des sources, dans le cas extrême où tous les éditeurs se soucieraient exclusivement de leur référencement sur Google, le moteur de recherche n'aurait plus d'intérêt véritable dans la mesure où les sujets traités seraient les mêmes.

Par ailleurs, Susan Gerhart a montré comment les moteurs de recherche avaient tendance à atténuer les controverses en privilégiant le bon côté des débats : « *Les moteurs de recherche, a-t-elle expliqué, ne conspirent pas dans le but de supprimer les controverses, mais leurs stratégies conduisent à la prédominance de sites organisationnels, ce qui prive les personnes en quête d'informations d'une expérience plus riche et, parfois, d'informations essentielles à la prise de décision* »<sup>9</sup>. Ainsi, pour une requête comportant le nom de pays « Belize », Google aura tendance à aiguiller l'internaute vers des pages présentant les caractéristiques géologiques, démographiques et touristiques du pays (qui ne font l'objet d'aucun débat), plutôt que de faire des liens vers des sites mentionnant le conflit frontalier du Belize avec le Guatemala. Les journalistes qui veulent optimiser leurs chances de figurer sur Google sont donc encouragés à privilégier les points de vue les plus consensuels.

En 2005, Eric Ulken a quant à lui analysé les réponses données par Google concernant Georges W. Bush et John Kerry, respectivement candidats du Parti républicain et du Parti démocrate lors de la campagne pour l'élection présidentielle américaine de 2004<sup>10</sup>. Ses conclusions montrent que Google pourrait avoir un biais politique, sans qu'on puisse dire à l'avance quel candidat est avantagé par ce biais. Il a également montré que le moteur Google avait tendance à privilégier les traitements négatifs de la campagne présidentielle. Ainsi, Google peut encourager les journalistes à prendre parti politiquement (sans que le bord pour lequel ils se prononcent soit identifiable).

### **Quand Google encourage les fautes journalistiques**

Dans un article du Poynter<sup>11</sup>, la journaliste Kelly McBride a révélé quelles pouvaient être les dérives de l'influence de Google sur le choix, non plus des sujets traités, mais des mots employés par les journalistes. Elle a montré comment le traitement journalistique du

projet de construction d'un centre musulman à deux « blocs » de maison du Ground Zero à New York avait pu être erroné. Puisque les internautes formulaient la requête « *ground zero mosque* » sur le moteur Google, de nombreux journalistes utilisaient ce terme à des fins de captation du trafic ; mais l'information donnée par leurs titres sous-entendait qu'une mosquée avait été construite sur le *Ground Zero*, ce qui était faux puisque la mosquée devait être construite beaucoup plus loin (les deux « blocs » de maison étant de taille considérable) et que le projet, porté par des musulmans modérés, n'était passeulement de construire une mosquée, mais également un centre ouvert à tous, séparé du lieu de culte et comprenant une piscine, un terrain de basket, une salle de gym, un auditorium de 500 places, une librairie, un studio d'art, un restaurant et une école de cuisine.

Ce point indique la limite de la stratégie qui consiste à reprendre dans le titre d'un article les termes utilisés par les internautes lors de leurs requêtes, et qui est pourtant plébiscitée par certains journalistes<sup>12</sup>. En effet, l'objectif de la requête est d'exprimer un besoin d'information, tandis que le but d'un article est de traiter l'information. Le premier est une question, le deuxième une réponse. Dès lors, la stratégie qui consiste à reprendre dans la réponse les termes utilisés dans la question, parfaitement valable du point de vue du SEO, peut conduire à certaines erreurs inacceptables du point de vue journalistique. Le lexique de la question, en influençant celui de la réponse, rend celle-ci inexacte. C'est ce que montre cet exemple très largement commenté en 2010 par les spécialistes de la presse en ligne américaine.

### **La légitimité et la déontologie des journalistes**

Finalement, les journalistes se retrouvent confrontés à des pressions émanant d'eux-mêmes davantage que de leur entourage, car même si personne ne les force directement à adhérer pleinement aux stratégies de référencement et à se livrer à des stratégies de SEO intensives au point de diminuer la qualité de leur travail, ils savent qu'il en va de la survie financière de leur entreprise et de leur légitimité professionnelle. Dans un espace où leur production côtoie celles des amateurs et des communicants travaillant pour le compte d'entreprises ou de partis politiques, les journalistes ont en effet besoin d'être visibles. Autrement dit, et selon l'expression consacrée, ils doivent « écrire pour être lus ».

Parce qu'ils jouissent en général d'une marge d'indépendance qui leur permet de décider de leur degré d'implication dans la stratégie de SEO, les journalistes se fixent eux-mêmes une limite à ne pas dépasser en la matière. Il peut alors arriver qu'ils soient partagés, entre eux, plutôt qu'opposés à la direction, au référenceur, aux équipes commerciales ou aux techniciens. En effet, les journalistes ne sont pas toujours

d'accord sur ce qui doit être fait en matière de SEO, et alors la déontologie, en plus d'être questionnée individuellement, est négociée entre eux : dans la salle de rédaction. Il serait donc inexact de dire qu'à cause d'enjeux liés au référencement, la déontologie des journalistes leur « échappe ».

### **Les référenceurs ne se mêleront pas du journalisme, si les journalistes apprennent le référencement**

De la littérature récente consacrée à la question, il ressort que l'action de Google est susceptible d'influencer l'ensemble du processus de production, de communication et de valorisation de l'information journalistique. Mais l'action de Google est aussi influencée par les actions de nombreux acteurs, qui pourront se conformer aux conditions et aux modalités de captation définies par les ingénieurs de Google tout en demeurant libres de les ignorer, de les détourner ou de les contourner. En définitive, l'influence est mutuelle et chaque acteur concerné par le fonctionnement du moteur Google peut jouer un rôle qui ne sera pas forcément celui qui a été prévu ou souhaité par les autres, et influencer ainsi, non seulement le résultat du processus de communication, mais aussi la nature même de ce processus, tout en influençant les actions, les possibilités d'action et les résultats des actions des autres.

Les journalistes, parce qu'ils sont tiraillés entre des considérations économiques et des objectifs éditoriaux, et parce qu'ils demeurent libres, ont tout intérêt à chercher à comprendre comment fonctionne Google et comment ils peuvent optimiser leur visibilité sur le moteur sans faire de concessions quant à ce qu'ils considèrent comme étant l'essence de leur rôle dans la société. Pour cela, il est important qu'ils apprennent certaines notions de code HTML qui leur permettront de discuter avec le référenceur et d'effectuer certaines opérations dont les résultats seront invisibles pour l'internaute et n'affecteront aucunement l'aspect éditorial de la production.

Aujourd'hui, les possibilités du code HTML sont trop souvent méconnues par les journalistes, qui se retrouvent obligés de traiter certains sujets en particulier, d'employer certains mots, d'uniformiser leurs titres ou de restreindre le nombre de liens hypertextes pointant vers des sources extérieures. S'ils maîtrisent les outils du référenceur, ils pourront renforcer la barrière déontologique au-delà de laquelle ce dernier n'aura pas accès, et négocier avec lui de « son côté » de la barrière. Il ne faut donc pas donner aux journalistes des formations d'écriture pour Google, comme c'est souvent le cas, mais des formations d'utilisation du code HTML pour optimiser son référencement sur le moteur sans avoir, justement, à modifier son écriture.

Sources :

1. « L'événement monstre », Pierre Nora, Communications, n° 18, 1972, p. 162-172. [Retour](#)
2. « La métamorphose de l'information : de sa production à sa coproduction », Rémy Rieffel, in Greffe X., Nathalie Sonnac (dir.), Culture Web, Paris, Dalloz, 2008, p. 103 [Retour](#)
3. « The Anatomy of a Large-Scale Hypertextual Web Search Engine » Computer Networks and ISDN Systems 30:107-117, Sergey Brin and Lawrence Page, 1998. [Retour](#)
4. « Method for Node Ranking in a Hyperlinked Database. U.S. Patent 6,285,999, filed January 9, 1998, and issued September 4, 2001 », Lawrence Page, 2001. [Retour](#)
5. Voir à ce sujet, « Dans l'esprit du PageRank. Une enquête sur l'algorithme de Google », Dominique Cardon, Réseaux, 1, n° 177, 2013, p. 63-95, et « What is in PageRank ? A Historical and Conceptual Investigation of a Recursive Status Index », Bernhard Rieder, Computational Culture, 28 septembre 2012, [http://computationalculture.net/article/what\\_is\\_in\\_pagerank](http://computationalculture.net/article/what_is_in_pagerank). [Retour](#)
6. « Search engine optimisation in UK news production », Murray Dick, Journalism Practice, vol. 5, n° 4, p. 462-477, 2011. [Retour](#)
7. Voir par exemple « Diversité et concentration de l'information sur le web. Une analyse à grande échelle des sites d'actualité français », Emmanuel Marty, Franck Rebillard, Stéphanie Pouchot, Thierry Lafouge Réseaux, vol. 6, 2012, n° 176, p. 27-72. [Retour](#)
8. « "Googlearchy": How a Few Heavily Linked Sites Dominate Politics on the Web », M. Hindman, K. Tsioutsoulis, J.A. Johnson Annual meeting of the Midwest Political Science Association (Chicago), 2003, 33 pages. [Retour](#)
9. « Do Web search engines suppress controversy ? », Susan Gerhart, First Monday, vol. 9, n° 1, 2004, [http://firstmonday.org/issues/issue9\\_1/gerhart/index.html](http://firstmonday.org/issues/issue9_1/gerhart/index.html) [Retour](#)
10. « Non-Traditional Sources Cloud Google News Results », E. Ulken, Online Journalism Review, 2005, <http://www.ojr.org/050519ulken/> [Retour](#)
11. « SEO Makes It Too Late for Truth for "Ground Zero Mosque" », K. McBride, Poynter, 2011, <http://www.poynter.org/latest-news/making-sense-of-news/105201/seo-makes-it-too-late-for-truth-for-ground-zero-mosque/> [Retour](#)
12. Voir par exemple « How SEO Is Changing Journalism », S Richmond, British Journalism Review, 19, 2008, p. 51-55 ou « Student Journalists Need to Learn SEO More Than They Need AP Style », R. Niles, Online Journalism Review, 2010. [Retour](#)