



Place des moteurs de recherche dans l'éditorialisation du web

Brigitte Simonnot

► To cite this version:

Brigitte Simonnot. Place des moteurs de recherche dans l'éditorialisation du web. *Communication & langages, Nec Plus*, 2016, 188, pp.45-59. 10.4074/S0336150016012035 . hal-01286514

HAL Id: hal-01286514

<https://hal.archives-ouvertes.fr/hal-01286514>

Submitted on 30 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Place des moteurs de recherche dans l'éditorialisation du web.

Droit à l'oubli contre droit à l'information

Brigitte Simonnot

Université de Lorraine

Centre de recherche sur les médiations (Crem EA3476)

brigitte.simonnot@univ-lorraine.fr

Biographie : Brigitte Simonnot est professeur en sciences de l'information et de la communication à l'université de Lorraine, France, et membre du Centre de recherche sur les médiations (EA 3476). Ses recherches portent notamment sur les dispositifs d'accès à l'information en ligne et les pratiques informationnelles. Sur cette thématique, elle a dirigé un dossier paru en 2008 dans la revue *Questions de communication*, et codirigé avec Gabriel Gallezot l'ouvrage *L'entonnoir. Google sous la loupe des sciences de l'information et de la communication*, paru en 2009 chez C&F éditions. Elle est aussi l'auteur de l'ouvrage *L'accès à l'information en ligne, moteurs, dispositifs et médiations* paru en 2012 chez Hermès science publications - Lavoisier.

Résumé : En vingt ans, les moteurs de recherche web ont conquis une place considérable dans le paysage informationnel. Face aux contestations de leurs politiques éditoriales, ils se sont longtemps abrités derrière une neutralité voire une objectivité, supposée résulter de traitements informatisés. De multiples controverses les ont poussés à clarifier leurs politiques. Qu'en disent les consignes données par les moteurs aux webmestres ? Dans celles de Google et de Bing, on peut trouver quelques points communs mais aussi des positionnements différents. Le référencement auprès des moteurs est devenu une activité à part entière et l'on s'est moins intéressé aux cas de déréférencement. La décision de la Cour de justice de l'Union européenne en 2014, dite du droit à l'oubli, en fournit l'occasion. Les réponses apportées par les moteurs soulignent certains aspects de leurs politiques éditoriales et invitent aussi à reconceptualiser la notion d'éditorialisation.

Mots-clés : moteur de recherche, consignes aux webmestres, neutralité, responsabilité éditoriale.

Title : Place of business search engines in online editorialization. Right to be forgotten against right to information

Abstract : For twenty years, web search engines have gained a considerable place in the information landscape. Facing protests against their editorial policies, they sheltered for a long time behind neutrality, even objectivity, supposedly resulting from a computerized processing. Many controversies push them to clarify their policies. What do instructions that they give to webmasters tell about these policies? In Google and Bing instructions, we can find some similarities but also some differences in their positioning. Search engine optimization has become a full-fledged activity while less interest has been brought to delinking. The ruling of the European court of justice in 2014, so-called “right to be forgotten”, gives an opportunity to take a look at it. The answers given by search engines to the ruling underline some aspects of their editorial policies. All this together invites to reconceptualize the notion of editorialization.

Keywords: search engine, webmasters guidelines, neutrality, editorial responsibility

1. Un paysage informationnel transformé par les moteurs

Depuis la fin des années 1990, l'accès à l'information s'est transformé de manière perceptible pour les internautes. C'est un lieu commun que de rappeler combien le développement et les évolutions du web ont contribué à l'accroissement de l'offre de contenus en ligne. Toutefois, ce sont les dispositifs d'accès à l'information en ligne, et en particulier les moteurs de recherche, qui ont contribué à l'installation d'une logique de la demande, en proposant des modalités d'accès fondées à la fois sur la singularité de l'internaute en quête de renseignement et sur les régularités des pratiques. Parmi les moteurs web aujourd'hui, on trouve à la fois des applications généralistes, qui donnent accès à toutes sortes et toutes formes de contenus, et des moteurs verticaux c'est-à-dire spécialisés dans un domaine ou un format de documents. Le paysage de ces portes d'entrée du web que sont les moteurs évolue très lentement en termes de diversité. En parts de marché, c'est toujours Google qui arrive en tête en Europe et en France (plus de 90% de parts calculées en nombres de requêtes), suivi de très loin par

Yahoo et Bing¹. Quelques nouveaux entrants tentent de s'installer : DuckDuckGo fondé en Pennsylvanie en 2008, ou le franco-allemand Qwant, lancé en 2013. DuckDuckGo affiche son intention d'évoluer vers un système de question/réponse c'est-à-dire de fournir directement l'information à l'internaute plutôt que de lui proposer des liens en résultat. Son infrastructure repose, outre son propre robot de collecte, sur une plateforme collaborative DuckDuckHack qui permet aux internautes de contribuer à produire des réponses synthétiques. Il exploite des sources « verticales », spécialisées dans certains domaines, et complète ses résultats avec ceux de Yahoo! et, dans d'autres régions, Yandex ou Bing². Qwant propose une version junior de son moteur. Ces deux entrants mettent en avant leur respect de la vie privée de l'internaute : les requêtes qui leur sont soumises ne sont pas « tracées ». Pour pouvoir répondre et rivaliser avec les moteurs plus anciens, ce sont non seulement leurs algorithmes qui importent mais surtout la taille de leurs index. C'est la raison pour laquelle, à leurs débuts, ces nouveaux entrants travaillent en collaboration avec des sources spécifiques voire d'autres moteurs. Les productions collaboratives en ligne contribuent à alimenter les index et les algorithmes des moteurs. On se souvient que Google s'était appuyé sur l'annuaire libre Dmoz pour constituer son index de démarrage³. L'encyclopédie collaborative Wikipedia est l'une des sources les plus exploitées pour construire, par exemple, des réponses synthétiques qui sont directement présentées dans les pages de résultats, à la manière du *knowledge graph* de Google.

Dans l'écologie de l'information⁴ en ligne, les moteurs occupent une place particulière. Sortes de méta-éditeurs, ils éditorialisent des contenus conçus et publiés par d'autres. L'objet de cet article est d'examiner les particularités de l'éditorialisation opérée par les moteurs web Bing et Google (Yahoo Search utilisant la technologie de Bing, son intervention se limite à la mise en page de résultats fournis par cette technologie). Pour ce faire, nous avons procédé à une analyse des consignes qu'ils donnent aux webmestres pour que les sites qu'ils conçoivent soient traités par leurs services. Ensuite, nous avons examiné comment les moteurs ont réagi à

¹ « Selon AT Internet, au mois de décembre 2014, 93,5% des requêtes effectuées sur des moteurs de recherche en France sont passées par Google. Suit Yahoo (2,6%), Bing (2,4%), Orange (0,4%) et Ask (0,3%). » Source : « Parts de marché des moteurs de recherche en France », *Journal du Net*, 25/07/2015, Accès : <http://www.journaldunet.com/ebusiness/le-net/parts-de-marche-des-moteurs-de-recherche-en-france.shtml> (consulté le 18/02/2016)

² « Sources », *Aide de DuckDuckGo*, <https://duck.co/help/results/sources>, consulté le 2/01/2016

³ Voir par exemple Gautam Pant, Padmini Srinivasan, Filippo Menczer, « Crawling the Web », In M. Levene & A. Poullovassilis (eds) *Web Dynamics. Adapting to Change in Content, Size, Topology and Use*. Berlin/Heidelberg, Springer Verlag, 2004, p. 153-177

⁴ L'expression d'écologie de l'information désigne ici les relations qui s'établissent entre des individus ou des groupes et le milieu informationnel dans lequel prennent place leurs pratiques, et les influences réciproques mais parfois asymétriques qui se jouent dans différentes situations.

la décision de la Cour de justice de l'union européenne dite du « droit à l'oubli ». Cette analyse met en lumière les particularités des politiques d'éditorialisation des deux moteurs commerciaux et invite à reconceptualiser la notion d'éditorialisation.

2. L'éditorialisation par les moteurs de recherche

Si l'activité traditionnelle d'édition consiste en la préparation de textes ou de contenus, en leur publication et en leur diffusion, l'éditorialisation de contenus numériques peut être définie comme la manière dont ces derniers sont sélectionnés, collectés et mis en page pour leur présentation. Sur le web, l'éditorialisation suppose la composition de pages réunissant un ensemble de (fragments de) textes, voire d'images fixes ou animées, et de liens menant à d'autres contenus. Bruno Bachimont⁵ retenait le qualificatif d'éditorialisation pour désigner cet « enrôlement » de ressources et leur composition dans une nouvelle publication, en soulignant le passage de technologies traditionnelles d'indexation de contenus en vue de recherches documentaires à des technologies d'indexation de fragments épars réunis sur un même sujet. Dans ce processus, il distinguait trois « postures » (*sic*) quant à la manière de tenir compte de l'origine des ressources : la posture amnésique, qui consiste à oublier le contexte original du fragment pour se focaliser uniquement sur son contenu (il donne l'exemple d'une photographie reprise pour illustrer un événement totalement différent de celui qu'elle immortalisait) ; la posture généalogique qui cherche à mettre en valeur les ressources dans leur contexte documentaire d'origine, à la manière des institutions patrimoniales ; et la posture créative qui réutilise des ressources en tant qu'objets matériels indépendamment de leur signification et de leur contenu perceptif, comme le font certaines initiatives d'art numérique. La présentation – contextualisée ou décontextualisée – des éléments sélectionnés est censée contribuer à la construction de sens chez le « lecteur ».

Les moteurs de recherche web, en sélectionnant et en classant des pages tirées de sites tiers, assurent en quelques sortes le rôle de méta-éditeurs : s'ils ne sont pas à l'origine de la publication des contenus qu'ils présentent et n'en contrôlent pas la mise en forme originale, ce sont eux qui leur donnent une visibilité et les rendent accessibles aux internautes. Cette accessibilité conditionne leur réutilisation potentielle. La manière dont ils sélectionnent les contenus et mettent en page les résultats correspond à une forme d'éditorialisation, qu'il

⁵ Bruno Bachimont, « Nouvelles tendances applicatives : de l'indexation à l'éditorialisation », In P. Gros (ed.) : *L'indexation multimédia: description et recherche automatiques*. Paris, Lavoisier, Hermès sciences, 2007, p. 313-326

s'agisse de répondre à la requête d'un internaute ou de proposer une page d'accueil à un service, par exemple de recherche d'actualités. Les formes données à cette éditorialisation varient en fonction des contenus sélectionnés. Pour le web « généraliste », la présentation des résultats par les moteurs web a connu une certaine standardisation : proposés sous forme d'une liste de liens accompagnés de courts descriptifs, ils sont ordonnés selon des critères inscrits dans les algorithmes. Ces critères portent sur la correspondance entre les termes de la requête de l'internaute et ceux des pages indexées, la position de ces termes dans les pages mais aussi, dans le cas du Pagerank de Google, sur des recommandations. Ces dernières sont opérationnalisées par un calcul du nombre de liens posés vers les pages par d'autres sites, pondéré par la notoriété des citants. Pour les moteurs d'actualités, ce sont la « fraîcheur » et la fréquence des sujets dans les titres de presse moissonnés qui importent davantage : les résultats regroupent sur un même sujet plusieurs articles tirés de titres de presse différents.

Dans le monde traditionnel de l'édition, l'éditeur se porte garant, après l'auteur, des contenus publiés. En effet, il met en place des processus de relecture qui permettent de corriger et valider ces contenus. Les moteurs de recherche web ne peuvent pas se porter garants des contenus qu'ils contribuent à promouvoir, à plus forte raison du fait que leur éditorialisation est automatisée. L'algorithmisation du processus est d'ailleurs un argument souvent invoqué pour justifier de leur « neutralité » quant au choix et à l'ordre de présentation des résultats qu'ils proposent. Les conditions générales d'utilisation de Google indiquent par exemple :

« Nos Services affichent des contenus n'appartenant pas à Google. Ces contenus relèvent de l'entière responsabilité de l'entité qui les a rendus disponibles. Nous pouvons être amenés à vérifier les contenus pour s'assurer de leur conformité à la loi ou à nos conditions d'utilisation. Nous nous réservons le droit de supprimer ou de refuser d'afficher tout contenu que nous estimons raisonnablement être en violation de la loi ou de notre règlement. Le fait que nous nous réservions ce droit ne signifie pas nécessairement que nous vérifions les contenus. Dès lors, veuillez ne pas présumer que nous vérifions les contenus. »⁶

Les critères qui permettent de sélectionner les contenus, de les ordonner et de les présenter évoluent au gré des mises à jour des algorithmes. Si les contenus indexés et promus

⁶ Conditions générales d'utilisation (CGU) de Google <http://www.google.fr/intl/fr/policies/terms/regional.html> .
Date de la dernière modification : 30 avril 2014 (consulté le 2/01/2016)

n'appartiennent pas aux moteurs, ces derniers en revendiquent le droit d'utilisation, de reproduction et de traduction⁷.

2.1 L'éditorialisation des pages de résultat

Dans la présentation des résultats des moteurs, l'intervention éditoriale se manifeste d'abord par l'organisation interne des pages. Ces dernières sont construites en trois ou quatre zones de contenus : la liste des résultats éditoriaux, les liens publicitaires, des suggestions de recherches associées et parfois un encadré. Les liens publicitaires s'affichent soit sur le côté de la page, soit en haut des résultats éditoriaux, parfois aux deux endroits. Depuis 2012, pour certaines requêtes, le *knowledge graph* de Google propose sur le côté de la page un encadré présentant un résumé informatif ; pour les noms communs ou les noms de personnalités, ces informations sont tirées de l'encyclopédie collaborative Wikipédia. Un procédé similaire est utilisé par le moteur Qwant (figure 1).

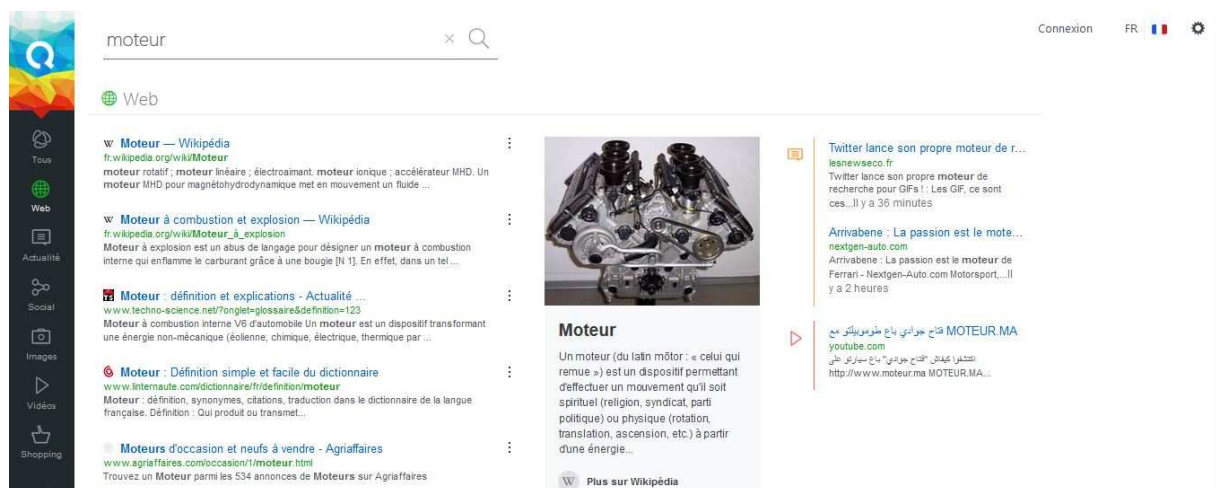


Figure 1 - Page de résultat du moteur Qwant proposant un encadré

Les résultats éditoriaux sont listés sous forme de paragraphes ordonnés. Pour chaque résultat est mentionné son titre, son adresse URL, un court résumé ou extrait et, éventuellement, une liste de rubriques spécifiques du site. Le texte du résumé, qui a été un temps composé par les moteurs pour mettre en valeur des termes de la requête tels qu'ils apparaissent dans la page, provient désormais soit de la balise méta-description du site, soit du début du texte de la page cible. Depuis peu, ces résumés peuvent présenter des « extraits enrichis », obtenus à partir de

⁷ « Lorsque vous importez, soumettez, stockez, envoyez ou recevez des contenus à ou à travers de nos Services, vous accordez à Google (et à toute personne travaillant avec Google) une licence, dans le monde entier, d'utilisation, d'hébergement, de stockage, de reproduction, de modification, de création d'œuvres dérivées (des traductions, des adaptations ou d'autres modifications destinées à améliorer le fonctionnement de vos contenus par le biais de nos Services), de communication, de publication, de représentation publique, d'affichage public ou de distribution publique desdits contenus. » CGU de Google, *ibid*.

données structurées encodées dans des microformats. Pour les produits ou les recettes de cuisine par exemple, le procédé permet d'inclure des critiques sous forme d'un classement par étoiles ou d'un nombre d'avis. Ce type de données requiert la participation des concepteurs de site pour spécifier dans le code de la page des balises associant des attributs à des propriétés, selon des standards ouverts⁸.

Loin de se contenter de collecter les contenus en ligne, les moteurs tentent également d'influencer la manière dont ils sont conçus et présentés, notamment à travers les consignes données aux webmestres dont ils requièrent la collaboration.

2.2 Consignes des moteurs à l'intention des webmestres

Bing et Google publient des consignes aux webmestres⁹ pour améliorer le référencement de leurs sites. L'analyse de ces consignes met en évidence quelques éléments communs mais aussi des spécificités. Y sont abordés le contenu des sites, les aspects techniques de la conception et une liste de « choses à éviter » (Bing) ou de « consignes relatives à la qualité » (Google). De manière paradoxale, les moteurs cherchent à dissuader les webmestres de concevoir leurs sites de manière spécifique pour qu'ils apparaissent bien classés dans les résultats mais les consignes qu'ils leur donnent les renvoient sans cesse aux exigences du moteur.

La cible des discours de Google est bien plus large que celle visée par Microsoft. En règle générale, les consignes de Google s'adressent d'abord aux responsables de sites (ceux qui décident de concevoir un site pour leurs « clients ») et non aux spécialistes techniciens auxquels la conception web est confiée ou déléguée. Ces responsables sont intimés de veiller à ce que les développeurs respectent les règles. Bing s'adresse plutôt aux développeurs en évoquant d'emblée des détails techniques. Ces détails, dans le cas de Google, sont relégués dans des pages dédiées signalées par des liens.

Bing fournit des consignes en anglais et une option de traduction automatique qui en donne une version en français approximatif (le lecteur est d'ailleurs invité à aider à améliorer cette traduction). L'introduction avertit que ces consignes ne sont pas exhaustives et invite le lecteur à consulter des documents d'aide, à suivre un blog dédié aux webmestres et à recourir aux outils d'analyse de référencement (sous réserve d'authentification). Google propose des

⁸ Voir le wiki dédié : <http://microformats.org/>

⁹ Bing : Mode d'emploi et aide webmaster, <https://www.bing.com/webmaster/help/help-center-661b2d18> ; Google : Search Console, <https://www.google.com/webmasters/> (consultés le 2/01/2016)

consignes en français ainsi qu'une formation à la conception de site web (dans la « Google Academy »). En commun, les deux fournisseurs de moteur insistent pour que les webmasters fournissent du contenu de qualité, « utile et riche en informations », organisé selon une structuration hiérarchique claire (Google), « clair et facile à trouver » (Bing). Les consignes techniques portent sur la manière de faciliter le travail des robots d'exploration et d'indexation, et conseillent de veiller à ce que le temps de chargement des pages soit minimal. Les « choses à éviter » (Bing) ou les « consignes relatives à la qualité » (Google) mettent l'accent sur les pratiques de conception de sites qualifiées d'abusives, qu'il s'agisse de mettre en œuvre des « combines » (Bing) ou des « astuces » (Google) pour améliorer de manière artificielle la visibilité du site. Les webmasters sont mis en garde sur le fait d'intégrer à leurs pages des contenus destinés uniquement aux moteurs, par exemple des mots ou des expressions populaires dans les recherches des internautes mais sans relation avec le contenu du site. Dans le même temps, les webmasters sont invités à consulter les statistiques des moteurs en termes de tendances de recherche¹⁰, c'est-à-dire les pages qui présentent périodiquement une synthèse statistique des requêtes reçues.

Les consignes insistent également sur la nature des liens hypertextes posés dans les pages : les webmasters ne doivent pas participer à des « systèmes de liens » concertés (*Link Schemes*), qu'il s'agisse de liens entrants ou sortants. Parmi les exemples de pratiques réprochées donnés par Google, figurent : l'achat ou la vente de liens destinés à promouvoir le site, y compris contre l'obtention de produits gratuits ; les campagnes de marketing à grande échelle recourant à des articles ou billets de blogs ; les publiereportages. En revanche, les liens publicitaires relevant du paiement au clic ne sont pas pénalisants à condition qu'ils soient exclus de l'exploration des robots par des directives spécifiques (via le recours à un attribut *no follow* ou une directive d'exclusion de la page dans le fichier robot.txt). Google invite les webmasters à créer des contenus de qualité qui seront remarqués par d'autres sites et susciteront des « votes éditoriaux » (*sic*) délibérés¹¹. Les webmasters sont ainsi enjoins de ne pas mettre en place des alliances entre eux pour influencer le référencement de leurs sites.

¹⁰ Tendances de Google : <https://www.google.fr/trends/> ; tendances de Bing : <http://www.bing.com/trends/> (consultées le 27/12/2015)

¹¹ : « La création d'un contenu de qualité est récompensée. Les liens constituent généralement des votes éditoriaux attribués de plein gré. Plus le contenu de votre site est utile, plus vous avez de chances qu'une autre personne trouve ce contenu intéressant pour ses lecteurs et crée un lien vers votre site. » Version en anglais : «The best way to get other sites to create high-quality, relevant links to yours is to create unique, relevant content that can naturally gain popularity in the Internet community. Creating good content pays off: Links are usually editorial votes given by choice, and the more useful content you have, the greater the chances someone else will

Contrairement à Google, Bing évoque de manière explicite les médias sociaux (rubrique intitulée « Social ») : « si vous êtes influent socialement, cela conduit les personnes qui vous suivent à partager largement votre information, ce qui permet à Bing de voir ces signaux positifs. Ces signaux positifs peuvent avoir un impact sur votre classement éditorial à long terme ». Une manière d'affirmer que le moteur prend en compte de ce type de sites mais aussi de relativiser leur impact qui ne peut être, si on en croit cette déclaration, qu'à long terme.

2.3 Surveiller et punir les pratiques jugées abusives

Dans les consignes aux webmestres, le ton des mises en garde contre les pratiques « abusives » est sensiblement différent chez les deux fournisseurs de moteur. Si Bing se contente d'indiquer que ces pratiques peuvent conduire à l'exclusion du site de l'index du moteur, Google évoque des mesures répressives graduées : « Elles [Les consignes] abordent en effet certaines des pratiques illicites susceptibles d'entraîner l'application d'une action manuelle pour cause de spam, d'une sanction de type algorithmique ou le retrait définitif des sites concernés de l'index Google »¹². Les conséquences annoncées sont non seulement le retrait du site des résultats éditoriaux de Google mais aussi de ses sites partenaires. La rubrique invite d'ailleurs le lecteur à signaler tout site suspecté de spam ou de pratiques de référencement abusives, via un lien qui mène à un formulaire dédié. Ces demandes et signalements ne sont pas traités au cas par cas mais de manière automatisée. Le texte précise également que l'intervention du moteur suite à ces signalements n'est pas toujours visible « au premier coup d'œil ».

Dans le monde des publications en ligne, la visibilité est essentielle. Les webmestres confient aux moteurs de recherche une partie de la mise en visibilité de leurs productions et certains jouent avec les paramètres supposés connus pour tenter de l'améliorer. Les moteurs, quant à eux, délèguent aux webmestres le soin de concevoir des pages qui soient adaptées au traitement des moteurs, tout en encadrant ces pratiques. Cependant ici, pas de contrat en bonne et due forme et aucune obligation de résultat, d'un côté comme de l'autre. Les fournisseurs de moteurs procurent bien à ceux qui le souhaitent des données statistiques sur la fréquentation de leur site (via par exemple Google Analytics), mais c'est peut-être un moyen supplémentaire pour obtenir des données à intégrer dans algorithmes.

find that content valuable to their readers and link to it.” Google search console help : <https://support.google.com/webmasters/answer/66356?hl=en> (consulté le 18/02/2016)

¹² Google Consignes aux webmasters - Aide Search Console. Accès : <https://support.google.com/webmasters/answer/35769?hl=fr> (consulté le 19/02/2016)

A travers leurs consignes, les fournisseurs de moteur cherchent à limiter les conséquences de leur intervention sur les contenus web qu'ils traitent et qu'ils exploitent. Ils entendent bien garder le contrôle éditorial de leurs pages de résultat, même si leurs architectures sont sensibles aux interventions de tiers.

3. Délégation de pouvoir en cascade

Déléguer, c'est charger quelqu'un d'une mission, généralement officielle, avec tout pouvoir pour la remplir ; ou encore transmettre, confier quelque chose, généralement des attributions officielles, à quelqu'un¹³. Bernard Rieder a été l'un des premiers à mobiliser le concept de délégation pour analyser ce qu'il nomme des « métatechnologies », c'est-à-dire des dispositifs technologiques « qui se bornent à structurer, filtrer, représenter, synthétiser et réguler les flux d'information et de communication »¹⁴. Il évoquait deux acceptions du terme, en politique et en gestion des organisations. Prenant appui sur la théorie de l'acteur-réseau¹⁵, il a montré la pertinence de ce concept pour analyser la manière dont se tissent dispositifs techniques et pratiques humaines, considérés dans leur contexte culturel et social. Le moteur vertical Google Actualités faisait partie des quatre études de cas qu'il a traitées.

Dès la création de Google, le fonctionnement du moteur a reposé sur une délégation de responsabilité particulière aux concepteurs de pages web. Tous les moteurs s'appuient sur les mots clés principaux présents dans une page pour l'indexer mais l'innovation de Google a été d'étendre ce procédé : les termes employés dans les liens qui mènent vers la page sont aussi intégrés à l'index. Google a donc été le premier à véritablement transformer les concepteurs de pages web en indexeurs, non seulement de leurs propres pages mais aussi de celles des autres. Ce principe – comparé par les concepteurs du moteur à un principe démocratique¹⁶ – revient à déléguer aux concepteurs de pages le pouvoir de décrire des pages dont ils ne sont pas les auteurs, avec des mots ou des expressions qui échappent aux auteurs. Bien avant l'avènement des modèles d'affaire du « web 2.0 », les concepteurs de Google avaient déjà compris le potentiel de la particularité réticulaire du web et du travail collaboratif. Là où le bât blesse, c'est que la firme californienne n'a pas informé clairement le public de cette disposition : cette dernière a été découverte par hasard par un journaliste formé à la recherche

¹⁴ Bernhard Rieder, *Métatechnologies et délégation. Pour un design orienté-société dans l'ère du Web 2.0*. Thèse de doctorat en SIC de l'université de Paris 8, 2006

¹⁵ Madeleine Akrich, Michel Callon et Bruno Latour (éd.), *Sociologie de la traduction : textes fondateurs*, Paris, Presses des Mines, 2006

¹⁶ Sergey Brin, Laurent Page, « The Anatomy of a Large-Scale Hypertextual Web Search Engine », *Computer Networks*, 30, 1998, p. 107-117.

d'information et à la documentation. Cette disposition, qui permettait une indexation et un classement original des pages web, s'est révélée être aussi une faille que les internautes pouvaient exploiter pour jouer sur le classement des résultats de certaines pages via ce que l'on nommait le *Google Bombing*¹⁷. Il devenait ainsi possible via des attaques groupées d'associer un site à des expressions ridicules ou malveillantes. En juillet 2005, Marissa Mayer, directrice des produits web consommateurs à Google (devenue PDG du concurrent Yahoo! en juillet 2012), indiquait sur le blog de la firme que cette dernière ne cautionnait pas ces pratiques mais qu'elle était réticente à modifier manuellement les résultats pour endiguer le phénomène. La porte-parole de l'entreprise réaffirmait à cette occasion que ces phénomènes « n'affectaient pas la qualité générale du moteur dont l'objectivité restait au cœur de leurs missions »¹⁸. C'est donc d'abord par des actions de communication – des communiqués sur sa bonne foi et insistant sur le fait qu'elle n'endossait pas les opinions promues par ces pratiques – que la firme a réagi avant de modifier son algorithme pour traiter le problème. Cette anecdote, parmi d'autres, montre combien la firme californienne recourt à la rhétorique de l'objectivité des algorithmes et rechigne à les remettre en cause pour répondre à des réclamations ou des détournements des utilisateurs.

3.1 La question du déréférencement

Si tout un secteur d'activité s'est créé autour des pratiques de référencement auprès des moteurs, les demandes de déréférencement méritent une attention particulière. En tant que fournisseurs de résultats, les moteurs web sont exposés à intégrer des contenus illicites, notamment des publications qui enfreignent la loi, par exemple le droit d'auteur, ou qui sont susceptibles de choquer certains publics. Quelles que soient leurs politiques éditoriales, ils doivent se conformer aux demandes de supprimer des accès à de tels contenus, dans le premier cas généralement suite à des décisions de justice, ou pour le second en proposant des filtres. Google communique sur les demandes de retrait de liens que la firme reçoit, qu'il s'agisse de demandes issues de l'administration, de sociétés gérant des droits d'auteur ou en Europe, ou de particuliers requérant le respect de leur vie privée. Aux États-Unis, le Digital Millennium Copyright Act (DMCA) prévoit depuis 1998 que les fournisseurs de moteurs ne peuvent être tenus responsables du fait qu'ils diffusent des contenus sous droits, à la condition expresse qu'ils s'engagent à supprimer ces liens dans un délai raisonnable.

¹⁷ Brigitte Simonnot, « Les médiations dans l'accès aux documents en ligne : pouvoir et autorité des moteurs commerciaux », *op. cit*

¹⁸ "Pranks like this may be distracting to some, but they don't affect the overall quality of our search service, whose objectivity, as always, remains the core of our mission."
<http://googleblog.blogspot.fr/2005/09/googlebombing-failure.html> (consulté le 25/02/2016)

Dans un autre registre, l'arrêt (C-131/12) du 13 mai 2014 de la Cour de justice de l'Union européenne (CJUE) instaurait un « droit à l'oubli », enjoignant les fournisseurs de moteurs de permettre aux internautes de demander la suppression de liens vers des contenus les concernant : « les personnes disposent d'un droit à demander le déréférencement d'informations en lien avec leur identité, sous réserve de certaines conditions et notamment de l'intérêt du public à avoir accès à l'information. [...] Cette suppression ne signifie pas l'effacement de l'information sur le site internet source. Le contenu original reste ainsi inchangé et est toujours accessible via les moteurs de recherche en utilisant d'autres mots clés de recherche ou en allant directement sur le site à l'origine de la diffusion. »¹⁹. Il s'agit donc de supprimer des index des moteurs, sur demande, certains documents mentionnant un nom de personne. L'arrêt de la CJUE met cependant en balance droit à la vie privée et droit à l'information : « Lorsque l'intérêt du public à être informé est prépondérant, le déréférencement n'est pas approprié ».

Les fournisseurs de moteurs ont dû mettre en place un formulaire de demande de déréférencement à l'intention des internautes européens et une procédure pour traiter ces requêtes. Google a proposé un formulaire dès mai 2014, suivi par Bing en juillet. Dans ses pages d'aide²⁰, Bing mentionne qu'il traite les demandes de retrait pour se conformer à la loi de certains pays, dans le cas de violation des droits d'auteur, de diffamation, d'informations permettant l'identification personnelle, de discours de haine ou pour respecter les autres droits de la personne mais il n'explique pas vraiment comment il le fait. Le moteur de Microsoft indique qu'il collabore avec des associations de régulation des contenus en ligne, notamment dans le domaine des abus contre les enfants. Pour les requêtes de déréférencement, il n'apporte aucune précision sur la manière dont il gère la procédure.

En 2014, Google a également publié un rapport sur « la transparence »²¹ où la firme rend compte du traitement des demandes de suppression de liens. Les statistiques sur les demandes traitées y sont actualisées. Le site donne des exemples rendus anonymes de demandes de retrait qui ont été satisfaites ou ignorées. Fin juillet 2014, Google disait avoir déjà reçu plus de

¹⁹ Commission nationale informatique et libertés, *Le droit au déréférencement en questions*, 21/09/2015, Accès : <https://www.cnil.fr/fr/le-droit-au-dereferencement-en-questions> (consulté le 19/02/2016)

²⁰ Bing, *How Bing delivers search results*, <http://help.bing.microsoft.com/#apex/18/en-US/10016/0>

²¹ Transparence des informations. *Google*. Accès : <http://www.google.com/transparencyreport/removals/europeprivacy/?hl=fr>. A noter que la *Wikimedia foundation en a fait de même* : <https://transparency.wikimedia.org/content.html>

91 000 demandes pour un total de 328 000 liens²², soit une moyenne d'environ 30 000 requêtes mensuelles. La firme a contribué à nourrir le débat sur le bienfondé de la décision de la CJUE. Au 25 février 2016, Google affichait 393 310 demandes nécessitant l'examen de 1 387 957 adresses URL²³, dont 57,5% ont été supprimées.

L'arrêt de la Cour n'oblige pas les moteurs de recherche à informer du déréférencement le site à l'origine de la diffusion du contenu, ni le public des suites données aux demandes de déréférencement. Cependant, Google a choisi de le faire. Lorsqu'une requête au moteur porte sur un nom de personne, un avertissement s'affiche en bas des pages de résultats : « Certains résultats peuvent avoir été supprimés conformément à la loi européenne sur la protection des données » (un lien suit vers la rubrique dédiée dans la foire aux questions du moteur). En outre, Google notifie ces suppressions aux sites à qui appartiennent les pages correspondantes. Un journaliste économique travaillant alors pour la BCC, Robert Peston, a mené l'enquête sur les raisons pour lesquelles un de ses billets de blog avait été retiré de l'index de Google. Dans la mesure où le moteur ne retire de son index que l'association entre le nom du demandeur et l'article, il est théoriquement possible de savoir qui en a fait la demande. En l'occurrence, dans le cas du billet du journaliste britannique, le litige portait sur un commentaire d'un internaute et non sur le billet lui-même. Les journalistes se sont émus de ce qu'un billet ou un article puisse être « censuré » des résultats du moteur non pour son contenu mais pour un commentaire d'internaute. BBC News publiait le 17 octobre 2014 un article déclarant que le média communiquerait la liste des adresses touchées par le « droit à l'oubli »²⁴ et a tenu sa promesse²⁵.

3.2 Le recours aux experts

Pour affiner sa politique suite aux demandes de retrait, et peut-être aussi pour asseoir sa légitimité, Google a réuni un comité d'experts en Europe²⁶. Ce conseil d'experts réuni à sept

²² « Wikipedia reveals Google 'forgotten' search links », *BBC News*, 6/08/2014, Accès : <http://www.bbc.com/news/technology-28672121> (consulté le 19/02/2016)

²³ « Demandes de suppression de contenus liés à la vie privée dans les résultats de recherche dans l'Union européenne », *Google Transparence des informations*, (consulté le 25/02/2016)

²⁴ Dave Lee, « BBC to publish 'right to be forgotten' removals list », *BBC News*, 17/10/2014, Accès : <http://www.bbc.com/news/technology-29658085> (consulté le 19/02/2016)

²⁵ Neil McIntosh, « Décembre 2015: List of BBC web pages which have been removed from Google's search results », *Internet blog BBC*, 04/01/2016. Accès : <http://www.bbc.co.uk/blogs/internet/entries/f4b01ccf-9128-45d8-8cac-23c1cf3455c1> (consulté le 22/02/2016)

²⁶ Le rapport final est publié en ligne : *The Advisory Council to Google on the Right to be Forgotten*, 6/02/2015 <https://drive.google.com/a/google.com/file/d/0B1UgZshetMd4cEI3SjlvV0hNbDA/view?pref=2&pli=1> (consulté le 19/02/2016)

reprises entre septembre et novembre 2014 en Europe s'est penché sur la question de savoir qui peut, de manière légitime, exercer un type de pouvoir sur l'information en ligne, sans porter atteinte à la liberté d'expression ou se livrer à une forme de censure. Le comité a retenu un certain nombre de critères permettant d'arbitrer les décisions. Tout d'abord, le rôle joué dans la vie publique par les demandeurs en constitue un premier : si une personne joue un rôle avéré dans la vie publique (par exemple un politicien, un chef religieux, une célébrité du spectacle ou du sport) ou locale (par exemple un directeur d'école), les informations qui la concerne sont sans doute d'intérêt public. Un second critère concerne la nature de l'information : la vie sexuelle, les revenus et finances, les données d'identité personnelle sont susceptibles d'être protégées. En revanche, les informations liées à un engagement politique, aux opinions religieuses ou philosophiques, à la santé publique et à la protection des consommateurs, ou encore à une activité criminelle sont d'intérêt public. Le comité recommande aussi de prendre en considération la source de l'information : par exemple, les publications d'un journaliste qui travaille dans les règles de l'art ou de personnes ayant une « bonne réputation » seront moins susceptibles de faire l'objet d'une suppression de liens. Enfin, le temps est également un critère qui permet d'évaluer si une information peut être oubliée, par exemple lorsqu'une personne a commis un délit mineur.

Le comité d'expert s'est également prononcé sur la procédure : la nature des données demandées dans le formulaire de requête de suppression de liens, et le fait d'avertir ou non l'éditeur de la page cible. Ses recommandations mentionnent notamment le fait qu'un éditeur puisse protester contre la suppression d'une de ses pages dans les résultats auprès d'une autorité comme la CNIL en France. Enfin, le comité s'est également penché sur la question territoriale : les liens déréférencés doivent-ils l'être seulement en Europe ou bien sur toutes les versions du moteur ? Face à la censure exercée dans certains pays, le comité a privilégié un déréférencement local.

On peut lire le rapport final produit par le comité comme une tentative de délimiter et de hiérarchiser des critères pour aider à la décision. Face au nombre important de requêtes en déréférencement, les firmes qui exploitent des moteurs cherchent à en automatiser le traitement.

3.3 Un pouvoir parajudiciaire attribué aux moteurs du web ?

Avec l'arrêt de 2014, on pourrait penser en première analyse que la CJUE a doté les fournisseurs de moteurs d'un pouvoir parajudiciaire : ce sont eux qui doivent apprécier si les demandes de retrait sont fondées ou non et le traitement qu'ils en font. Doter des compagnies

commerciales privées d'un tel pouvoir peut sembler abusif et absurde, puisque les contenus incriminés peuvent rester disponibles sur leur site de publication d'origine. Luciano Floridi, qui a fait partie du comité d'experts européens mis en place par Google, indiquait dans un article du Guardian que le plus grand nombre de demandes de déréférencement au moteur concernaient des pages Facebook. Dans de tels cas, écrivait-il, « il semble[rait] raisonnable de penser que la personne sollicite d'abord la suppression de cette information personnelle [sur le site d'origine], avant de demander qu'elle soit déréférencée »²⁷.

Cependant, cette décision de la CJUE peut aussi sembler justifiée pour plusieurs raisons liées à l'efficacité du dispositif, à la conscience de ce qu'est le web aujourd'hui et à la connaissance que les publics en ont. D'une part, à l'époque, les moteurs conservaient en mémoire (cache) une copie des contenus visés et les mettaient à disposition des internautes : les retirer du site d'origine n'aurait pas empêché de les trouver. D'autre part, sans les moteurs, de tels contenus deviennent pratiquement invisibles. Enfin, cette mesure peut aussi être interprétée comme une action visant à contraindre les moteurs à mieux respecter les législations en vigueur dans les différents pays. Elle est tout à la fois reconnaissance du pouvoir de ces dispositifs méta-éditoriaux dans l'environnement actuel de l'information en ligne et rappel à l'ordre quant aux responsabilités qui sont les leurs, responsabilités qu'ils ont tendance à nier dans leurs conditions générales d'utilisation.

4. Reconceptualiser la notion d'éditorialisation

Google a été le premier à implémenter un service à large échelle qui exploite les attachements entre sites en tenant compte des liens posés entre eux pour les classer. Cependant, le moteur veut rester maître de l'éditorialisation de ses pages de résultats. Or, la dynamique qui permet la production des résultats par les moteurs web repose sur un jeu de recommandations en spirale à plusieurs niveaux : pour être bien classés, les sites doivent être « recommandés » par des liens venant d'autres sites mais ils ne doivent pas recourir à des alliances trop visibles ou trop grossières. Les moteurs, quant à eux, tentent de donner des consignes et des recommandations aux webmestres, sans révéler le détail du fonctionnement de leurs algorithmes.

L'éditorialisation produite par l'algorithme Pagerank pour classer les résultats s'appuie largement sur des opinions, exprimées par les concepteurs de site à travers leurs choix en

²⁷ Luciano Floridi, « Right to be forgotten: who may exercise power, over which kind of information? », *The Guardian*, 21/10/2014 (consulté le 25/02/2016), notre traduction.

termes de liens sortants. Ce procédé a souvent été rapproché des techniques bibliométriques qui permettent de classer les publications scientifiques et les chercheurs, des techniques controversées. Marcello Vitali-Rosati²⁸ souligne les distinctions entre l'anglais *to editorialize*, « exprimer une opinion dans un éditorial », et le terme français au sens plus large. Le chercheur définit finalement l'éditorialisation non pas comme « la manière dont nous produisons des connaissances en utilisant des outils numériques [mais comme] la manière dont nous produisons des connaissances à l'âge du numérique ou, mieux, dans une société numérique », et plus précisément, comme « l'ensemble des dynamiques qui produisent et structurent l'espace numérique ». Pour ce chercheur, de la même manière que le langage est performatif, l'éditorialisation est un moyen de produire la réalité et non un moyen de la représenter. Dépenser le paradigme de la représentation pour adopter celui de la performativité permet d'analyser plus à fond la dynamique des publications en ligne et de leur mise en visibilité. Sur le web, l'éditorialisation prend des formes multiples fondées sur des valeurs culturelles, et elle relève d'actions collectives et de négociations.

4. Conclusion

Pendant longtemps, on a cru que les moteurs avaient un pouvoir réduit sur les contenus qu'ils éditorialisent dans leurs résultats. Elisabeth Van Couvering²⁹ par exemple considérait leur pouvoir limité dans la mesure où, à la différence d'autres industriels de la culture comme Vivendi, Disney, Time Warner ou Bertelsmann, ils ne créent pas de contenus et n'ont pas de contrôle effectif sur les contenus qu'ils contribuent à promouvoir. Toutefois, les moteurs contrôlent la distribution de ces contenus, comme le font dans d'autres registres Amazon ou des sites commerciaux comme e-Bay, et ce d'autant plus lorsqu'ils captent des parts de marché considérables dans certains pays. Il est également opportun de rappeler que la plupart des sociétés qui conçoivent des moteurs web se sont impliquées dans des projets de numérisation à grande échelle (livres ou archives de presse) et dans la conception de moteurs verticaux qui alimentent les résultats de leurs moteurs généralistes.

Les moteurs Bing et Google ont des politiques éditoriales et de communication très différentes. Le premier semble réduire son intervention à un simple outil technique. Le second, peut-être parce qu'il a été mis plus souvent en cause quant à sa transparence,

²⁸ Marcello Vitali-Rosati, « What is editorialization? », *Sens public*, 2016/01. Accès : <http://www.sens-public.org/article1059.html> (consulté le 4/01/2016)

²⁹ Elisabeth Van Couvering, « Navigational Media: The Political Economy of Online Traffic ». In D. Winseck and D. Y. Yin (eds.), *The Political Economies of Media. The Transformation of the Global Media Industries*, London, Bloomsbury, 2011.

communiqué de façon abondante et montre une conscience de la part de ses dirigeants des enjeux liés à leur activité. Certains commentateurs ont aussi soupçonné Google de tenter de soulever les journalistes contre l'arrêt dit du « droit à l'oubli », pour mettre l'opinion publique en émoi.

La décision de la Cour de justice de l'Union européenne de 2014 est venue rappeler le pouvoir effectif acquis par les moteurs. En même temps, elle a contribué à déplacer l'attention de la transparence des algorithmes vers la transparence de leurs prises de décision, fussent-elles « éclairées » par des experts. Elle a aussi conforté les moteurs dans leur rôle de porte-paroles en les enjoignant de concilier droit au caractère privé des données et liberté d'information. Dans ce domaine, les moteurs jouent encore un rôle très inégal et ambivalent. Il nous faudra suivre, dans la durée, la manière dont ces acteurs font évoluer la question et l'usage public qu'ils font de la raison³⁰, sans se contenter d'analyser la manière dont ils communiquent sur ce qu'ils font, même si ces communications contribuent à structurer le paysage numérique.

Références

Akrich Madeleine, Callon Michel et Latour Bruno (éds), *Sociologie de la traduction : textes fondateurs*, Paris, Presses des Mines, 2006

Bachimont Bruno, « Nouvelles tendances applicatives : de l'indexation à l'éditorialisation », In P. Gros (ed.) : *L'indexation multimédia: description et recherche automatiques*. Paris, Lavoisier, Hermès sciences, 2007, p. 313-326

Brin Sergey, Page Laurent, « The Anatomy of a Large-Scale Hypertextual Web Search Engine », *Computer Networks*, 30, 1998, p. 107-117

CNIL, *Le droit au déréférencement en questions*, 21/09/2015, Accès : <https://www.cnil.fr/fr/le-droit-au-dereferencement-en-questions>

Elgesem Dag, « Search engines and the public use of reason », *Ethics and Information Technology*, 10, 2008, p. 233–242

³⁰ Dag Elgesem, « Search engines and the public use of reason », *Ethics and Information Technology*, 10, 2008, p. 233–242.

Pant Gautam, Srinivasan Padmini, Menczer Filippo, « Crawling the Web », In M. Levene & A. Poulouvassilis (eds) *Web Dynamics. Adapting to Change in Content, Size, Topology and Use*. Berlin/Heidelberg, Springer Verlag, 2004, p. 153-177

Rieder Bernhard, *Métatechnologies et délégation. Pour un design orienté-société dans l'ère du Web 2.0*. Thèse de doctorat en SIC de l'université de Paris 8, 2006

Simonnot Brigitte, « Les médiations dans l'accès aux documents en ligne : pouvoir et autorité des moteurs commerciaux ». In : E. Broudoux et G. Chartron (eds.), *Enjeux politiques du document numérique, Actes de la troisième conférence Document numérique et Société*, Aix-en-Provence, ADBS Éditions, 2010, p.175-191

Van Couvering Elizabeth, « Navigational Media: The Political Economy of Online Traffic ». In D. Winseck and D. Y. Yin (eds.), *The Political Economies of Media. The Transformation of the Global Media Industries*, London, Bloomsbury, 2011

Vitali-Rosati Marcello, « What is editorialization? », *Sens public*, 2016/01. Accès : <http://www.sens-public.org/article1059.html>